

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 146.817 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-34
Perfect score: 7741
Sequence: 1 MDKKVRYKLRKVKGRVTVS.....DRYGRISYDANSGERVRIN 1475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*
9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7741	100.0	1475	5	Aau98027 S. mutans
2	7741	100.0	1475	7	Adx93654 Streptoco
3	7741	100.0	1475	9	Adx37277 Streptoco
4	7740	99.9	1475	5	Aau98030 S. mutans
5	7737	99.9	1475	5	Aau98040 S. mutans
6	7736	99.9	1475	5	Aau98031 S. mutans
7	7735	99.9	1475	5	Aau98033 S. mutans
8	7734	99.9	1475	5	Aau98032 S. mutans
9	7729	99.8	1475	5	Aau98035 S. mutans
10	7729	99.8	1475	5	Aau98034 S. mutans
11	7727	99.8	1475	5	Aau98036 S. mutans
12	7721	99.7	1475	5	Aau98037 S. mutans
13	7714	99.7	1475	5	Aau98039 S. mutans
14	7711	99.6	1475	5	Aau98038 S. mutans
15	7628.5	98.5	1476	5	Aau79284 Streptoco
16	5285	68.3	1375	5	Aau98028 S. mutans
17	5285	68.3	1375	5	Aau79288 Streptoco
18	5285	68.3	1375	5	Adx93655 Streptoco
19	5285	68.3	1375	9	Adx37278 Streptoco
20	5154.5	66.6	1017	5	Aau79285 Streptoco
21	4580	59.2	1590	7	Adx93657 Streptoco
22	4580	59.2	1590	9	Adx37280 Streptoco
23	4531	58.5	1592	2	Aar32925 Glucosylt
24	3829.5	49.5	1430	5	Aau98044 S. mutans

ALIGNMENTS

RESULT 1
AAU98027
ID AAU98027 standard; protein; 1475 AA.
XX AC AAU98027;
XX DT 27-AUG-2002 (first entry)
XX DE S. mutans glucosyltransferase GTFB.
XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture.
XX OS Streptococcus mutans.
XX PN US2002031826-A1.
XX PD 14-MAR-2002.
XX PF 19-DEC-2000; 2000US-00740274.
XX PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
(NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX N-PSDB; ABK52938.
PT Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
XX manufacture, comprises mutations in specific positions.
XX Disclosure; Page 21-25; 4pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

25 3825.5 49.4 1430 5 AAU98045 S. mutans
26 3824.5 49.4 1430 5 AAU98041 S. mutans
27 3822.5 49.4 1430 5 AAU98043 S. mutans
28 3820.5 49.4 1430 5 AAU98042 S. mutans
29 3817.5 49.3 1430 5 AAU98029 S. mutans
30 3817.5 49.3 1430 7 ADD93656 Streptoco
31 3815.5 49.3 1430 9 ADX37279 Streptoco
32 3695.5 47.7 1499 7 ADC54806 Protein S
33 3506 45.3 1577 2 AAR91047 Alpha-D-g
34 3434.5 44.4 1554 7 ADD93658 Streptoco
35 3434.5 44.4 1554 9 ADX37281 Streptoco
36 3398.5 43.9 2835 6 ABR98574 Dextran s
37 3398.5 43.9 2835 6 ABR55594 Amino aci
38 3215.5 41.5 1477 9 ADY72696 Mutant de
39 3214.5 41.5 1477 9 ADY72733 Mutant de
40 3210.5 41.5 1477 9 ADY72732 Mutant de
41 3197 41.3 1527 5 AAU80055 Leuconost
42 3196.5 41.3 1518 7 ADD93660 Streptoco
43 3196.5 41.3 1518 9 ADX37283 Streptoco
44 3195.5 41.3 1527 7 ADC54807 Leuconost
45 3161 40.8 1497 6 ABR63234 Glucanruc

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents GTFB
XX
SQ Sequence 1475 AA;

Query Match 100.0%; Score 7741; DB 5; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDKKRYKLRKVKRWTVSVASAVMTLTLSGGLVKADSNESKQIENDSNTSVVTANE 60
DB 1 MDKKRYKLRKVKRWTVSVASAVMTLTLSGGLVKADSNESKQIENDSNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVVSPPYVGETASNGEKLQNTT 120
DB 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVVSPPYVGETASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISQTTTEADTVIDDNSAANLQILEKLPNVKEIDGKYIYDNNQKVRT 180
DB 121 TVDKTSEAAANNISQTTTEADTVIDDNSAANLQILEKLPNVKEIDGKYIYDNNQKVRT 180
QY 181 NPTLIADGKILHFDGTGAYTDSIDTVAKDITVTRSNLYKKYNQVYDRSAQSFEHVDHYL 240
DB 181 NPTLIADGKILHFDGTGAYTDSIDTVAKDITVTRSNLYKKYNQVYDRSAQSFEHVDHYL 240
QY 241 TAESWYRPKYILKOGKWTQSTOKDFRPLMTWPDQETQROYVNMMAQLGINKTYDDT 300
DB 241 TAESWYRPKYILKOGKWTQSTOKDFRPLMTWPDQETQROYVNMMAQLGINKTYDDT 300
QY 301 SNQLQLNIAAATIQAIEAKITTLKNTDMLRQTIISAFVKTSAMNSDSEKPPDDHLONGA 360
DB 301 SNQLQLNIAAATIQAIEAKITTLKNTDMLRQTIISAFVKTSAMNSDSEKPPDDHLONGA 360
QY 361 VLYDNEGKLTYPANSNYRILNPTNQTGKDPRTYADNTTGGYEFLLANDVDNPNPVQ 420
DB 361 VLYDNEGKLTYPANSNYRILNPTNQTGKDPRTYADNTTGGYEFLLANDVDNPNPVQ 420
QY 421 AEQLNLWHLFNMFGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYKAAKGIHKNDKA 480
DB 421 AEQLNLWHLFNMFGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYKAAKGIHKNDKA 480
QY 481 ANDHLSILEAWSNDPTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMNPILTNSLV 540
DB 481 ANDHLSILEAWSNDPTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMNPILTNSLV 540

QY 541 NRTDDNAETAAPVPSYFIRAHDSVQDIIADIIKAEINPNVGVSYFTMEEIKKAEIYNK 600
DB 541 NRTDDNAETAAPVPSYFIRAHDSVQDIIADIIKAEINPNVGVSYFTMEEIKKAEIYNK 600
QY 601 DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGDMDTDDGOYMAHKNTINYEAIETLLKA 660
DB 601 DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGDMDTDDGOYMAHKNTINYEAIETLLKA 660
QY 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDGTDRTRTSGVAVIEGNNPSLRUK 720
DB 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDGTDRTRTSGVAVIEGNNPSLRUK 720
QY 721 ASDRVVNVNMGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIPTAADIKG 780
DB 721 ASDRVVNVNMGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIPTAADIKG 780
QY 781 YANPOVSGYLGWVPVGAALIKMFALRRLARPHQOMASVHQNAALDSRVMEFGFSNFQAPA 840
DB 781 YANPOVSGYLGWVPVGAALIKMFALRRLARPHQOMASVHQNAALDSRVMEFGFSNFQAPA 840
QY 841 TKKBEYTNVIAKNVDKFAEWGVTDFEMAPQVVSSTGSLDSVTQNGYAFTRDRLDGLIS 900
DB 841 TKKBEYTNVIAKNVDKFAEWGVTDFEMAPQVVSSTGSLDSVTQNGYAFTRDRLDGLIS 900
QY 901 KPNKYGTADDLVKAALKALHSKGIKVMADWVPDQMYAFPEKEVVATRVDKYGTPTVAGSQI 960
DB 901 KPNKYGTADDLVKAALKALHSKGIKVMADWVPDQMYAFPEKEVVATRVDKYGTPTVAGSQI 960
QY 961 KNTLYVVDGKSGKQQQAKYGGAFLEELQAKYPFELFARKQISTGVPMPPSVKIKOWSAKY 1020
DB 961 KNTLYVVDGKSGKQQQAKYGGAFLEELQAKYPFELFARKQISTGVPMPPSVKIKOWSAKY 1020
QY 1021 FNGNTILGRGAGYVLLKQATNTYFNI SDNKEINFLPKTLNQNDSQVSGSYDGKGYVYST 1080
DB 1021 FNGNTILGRGAGYVLLKQATNTYFNI SDNKEINFLPKTLNQNDSQVSGSYDGKGYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYVFDNNGYVMTCAQSTINGVNYFLSNGLQRLDAILKNEGDTYAY 1140
DB 1081 SGYQAKNTFISEGDKWYVFDNNGYVMTCAQSTINGVNYFLSNGLQRLDAILKNEGDTYAY 1140
QY 1141 YGNDGRRYENGYQFMGSGVWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFVTTADGK 1200
DB 1141 YGNDGRRYENGYQFMGSGVWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFVTTADGK 1200
QY 1201 IRYFDKQSGNMYRNFIEBEGKWLILGEDGAAVTGQTINGQHLVFRANGVQVKGEFVT 1260
DB 1201 IRYFDKQSGNMYRNFIEBEGKWLILGEDGAAVTGQTINGQHLVFRANGVQVKGEFVT 1260
QY 1261 DHGGRISYDNGSGDQIRNRFVRNAQGMWYFDNNGYAVTGARTINGQHLVFRANGVQVK 1320
DB 1261 DHGGRISYDNGSGDQIRNRFVRNAQGMWYFDNNGYAVTGARTINGQHLVFRANGVQVK 1320
QY 1321 GEFTVTDYGRISYDNGSGDQIRNRFVRNAQGMWYFDNNGYAVTGARTINGQHLVFRAN 1380
DB 1321 GEFTVTDYGRISYDNGSGDQIRNRFVRNAQGMWYFDNNGYAVTGARTINGQHLVFRAN 1380
QY 1381 GVQVKGEFVTDYGRISYDNGSGDQIRNRFVRNAQGMWYFDNNGYAVTGARTINGQHL 1440
DB 1381 GVQVKGEFVTDYGRISYDNGSGDQIRNRFVRNAQGMWYFDNNGYAVTGARTINGQHL 1440
QY 1441 YFRANGVQVKGEFVTDYGRISYDNGSGDQIRNRFVRNAQGMWYFDNNGYAVTGARTING 1475
DB 1441 YFRANGVQVKGEFVTDYGRISYDNGSGDQIRNRFVRNAQGMWYFDNNGYAVTGARTING 1475

RESULT 2
ADD93654
ID ADD93654 standard; protein; 1475 AA.
XX ADD93654;
XX
DT 29-JAN-2004 (first entry)

XX Streptococcus mutans glucosyltransferase-B.
DE Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
KW Streptococcus mutans.
XX WO2003075845-A2.
XX 18-SEP-2003.
XX 07-MAR-2003; 2003WO-US006962.
XX 07-MAR-2002; 2002US-0363209P.
XX 08-AUG-2002; 2002US-0402483P.
XX (FORS-) FORSYTH INST.
XX Smith DJ, Taubman MA;
XX WPI; 2003-845091/78.
XX Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 16; Page|12-13; 49pp; English.
XX The present sequence is the protein sequence of Streptococcus mutans
CC glucosyltransferase-B (Gtf-B). Peptide fragments of Gtf-B, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutans glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Diepitopic or
CC multi-epitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX Sequence 1475 AA;
SQ

Query Match 100.0%; Score 7741; DB 7; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDKVRKYLKRVKRWVTVSVASAVMTLTLSGGLVKADSNESKSI SNTSVVTANE 60
DB 1 MDKVRKYLKRVKRWVTVSVASAVMTLTLSGGLVKADSNESKSI SNTSVVTANE 60

QY 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSNPYTVGETASNGEKLQNTT 120
DB 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSNPYTVGETASNGEKLQNTT 120

QY 121 TVDKTSEAAANNISKOTTETADTDVIDDSNAANLQILEKLPNKEIDGKYIYDNNKQVET 180
DB 121 TVDKTSEAAANNISKOTTETADTDVIDDSNAANLQILEKLPNKEIDGKYIYDNNKQVET 180

QY 181 NFTLIADGKILHFDGTGYTDTSIDTVNKKDI VTTRSNLYKKYNQVYDRSAQSFEHVDHYL 240
DB 181 NFTLIADGKILHFDGTGYTDTSIDTVNKKDI VTTRSNLYKKYNQVYDRSAQSFEHVDHYL 240

QY 241 TAESWYRPKYLKDGKTWTQSTEKQDPRLLMTWPDQETQORQVNTMNAQLGINKTYDDT 300
DB 241 TAESWYRPKYLKDGKTWTQSTEKQDPRLLMTWPDQETQORQVNTMNAQLGINKTYDDT 300

QY 301 SNQLQNLIAAATIQAKIEAKITTLKNTDWLROTI SAFVKTQSAWNSDSEKPFDDHLQNGA 360
DB 301 SNQLQNLIAAATIQAKIEAKITTLKNTDWLROTI SAFVKTQSAWNSDSEKPFDDHLQNGA 360

QY 361 VLYDNEGKLTPTVANSNYRILNRTPTNQTGKDPRTADNTIGGYEFLLANDVDSNPVVQ 420

Db 361 VLYDNEGKLTPTVANSNYRILNRTPTNQTGKDPRTADNTIGGYEFLLANDVDSNPVVQ 420

QY 421 AEQLNWLHFLMFGNIYANDPDANFDSIRVDAVDNDADLLOIAGDYVLYKAAGHKNKDKA 480

Db 421 AEQLNWLHFLMFGNIYANDPDANFDSIRVDAVDNDADLLOIAGDYVLYKAAGHKNKDKA 480

QY 481 ANDHLSILEAWSNDTPYLLHDDGDNNIMDNKLRLLSLLFSLAKPLQSRQSMNPLITNSLV 540

Db 481 ANDHLSILEAWSNDTPYLLHDDGDNNIMDNKLRLLSLLFSLAKPLQSRQSMNPLITNSLV 540

QY 541 NRTDDNAETAAPVSYSFIRAHDSVQDLIADIIKABINPNVVGYSFTMBEIKKAFEIYNK 600

Db 541 NRTDDNAETAAPVSYSFIRAHDSVQDLIADIIKABINPNVVGYSFTMBEIKKAFEIYNK 600

QY 601 DLLLATEKKYTHYNTALSYALLTNKSSVPRVYVYGDMDTDDGQYMAHKTINYEAEITLKA 660

Db 601 DLLLATEKKYTHYNTALSYALLTNKSSVPRVYVYGDMDTDDGQYMAHKTINYEAEITLKA 660

QY 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKALKATDTGDRTRTTSVAVIEGNNPSLRK 720

Db 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKALKATDTGDRTRTTSVAVIEGNNPSLRK 720

QY 721 ASDRVVNNGAHKNQAYRPLLTITDNGIKAYHSDQEAAGLVRYTNDRGLIIFTAADIKG 780

Db 721 ASDRVVNNGAHKNQAYRPLLTITDNGIKAYHSDQEAAGLVRYTNDRGLIIFTAADIKG 780

QY 781 YANPOVSGYLVGVVPGVGAALIKMFALRLARPHQOMASVHQNAAALDSRVMEPEGSFOAPA 840

Db 781 YANPOVSGYLVGVVPGVGAALIKMFALRLARPHQOMASVHQNAAALDSRVMEPEGSFOAPA 840

QY 841 TKKEEYTNVVIKNDVKFAEWGVTDFEMAPQVYSSVSDGSLDSVIQNGVAFTRDYDLGIS 900

Db 841 TKKEEYTNVVIKNDVKFAEWGVTDFEMAPQVYSSVSDGSLDSVIQNGVAFTRDYDLGIS 900

QY 901 KKNKGTADDLKAIALKHSKGIKNWADVPQMYAFPEKEVVVATRVKIGTFPVAGSQI 960

Db 901 KKNKGTADDLKAIALKHSKGIKNWADVPQMYAFPEKEVVVATRVKIGTFPVAGSQI 960

QY 961 KNTLYVVDGKSGKQQAQYKGAFLLELOAKYPELFARKQISTGVPMDSVKIKOWSAKY 1020

Db 961 KNTLYVVDGKSGKQQAQYKGAFLLELOAKYPELFARKQISTGVPMDSVKIKOWSAKY 1020

QY 1021 FNGTNILGRGAGYVLKQDQATNTYFNISDNKEINFLPKILLNODSQQVFSYDGGKYVYYST 1080

Db 1021 FNGTNILGRGAGYVLKQDQATNTYFNISDNKEINFLPKILLNODSQQVFSYDGGKYVYYST 1080

QY 1081 SGYQAKNTFISBGDKWYIFDNNGYMVTGAQSIQNGVNYVFLSNGQLRLDAILKKNEDGTAY 1140

Db 1081 SGYQAKNTFISBGDKWYIFDNNGYMVTGAQSIQNGVNYVFLSNGQLRLDAILKKNEDGTAY 1140

QY 1141 YGNDGRRYENGYYQPMGSGVWRHFNNGEMSVGLTVIDGQVOYFDEMGOQAKGFVTTADGK 1200

Db 1141 YGNDGRRYENGYYQPMGSGVWRHFNNGEMSVGLTVIDGQVOYFDEMGOQAKGFVTTADGK 1200

QY 1201 IRYFKQSGNMVNRNFIENEKGKWLVLGSDGAATVGSQTINGOHLFRANGVOVKGFEVT 1260

Db 1201 IRYFKQSGNMVNRNFIENEKGKWLVLGSDGAATVGSQTINGOHLFRANGVOVKGFEVT 1260

QY 1261 DHHGRISYYDGNSSGDIQIRNRFVRNAQGWFFYFNNNGYAVTGARTINGQLLYFRANGVOVK 1320

Db 1261 DHHGRISYYDGNSSGDIQIRNRFVRNAQGWFFYFNNNGYAVTGARTINGQLLYFRANGVOVK 1320

QY 1321 GEFVTRDGRISYYDGNSSGDIQIRNRFVRNAQGWFFYFNNNGYAVTGARTINGOHL 1380

Db 1321 GEFVTRDGRISYYDGNSSGDIQIRNRFVRNAQGWFFYFNNNGYAVTGARTINGOHL 1380

QY 1381 GVOVKGFEVTDHGRISYYDGNSSGDIQIRNRFVRNAQGWFFYFNNNGYAVTGARTINGOHL 1440

Db 1381 GVOVKGFEVTDHGRISYYDGNSSGDIQIRNRFVRNAQGWFFYFNNNGYAVTGARTINGOHL 1440

QY 1441 YFRANGVOVKGFEVTDHGRISYYDGNSSGDIQIRNRFVRNAQGWFFYFNNNGYAVTGARTINGOHL 1475

Db 1441 YFRANGVOVKGFEVTDHGRISYYDGNSSGDIQIRNRFVRNAQGWFFYFNNNGYAVTGARTINGOHL 1475

RESULT 3

```

ADX37277
ID ADX37277 standard; protein; 1475 AA.
XX AC ADX37277;
XX DT 21-APR-2005 (first entry)
XX DE Streptococcus mutant glucan binding protein B variant #6.
XX KW immunogenicity; immune stimulation; glucan binding protein-B;
XX KW microparticle; major histocompatibility complex; tooth disease.
XX OS Streptococcus mutans.
XX PN US2005031633-A1.
XX PD 10-FEB-2005.
XX PF 09-MAR-2004; 2004US-00797821.
XX PR 13-APR-1998; 98US-0081550P.
XX PR 08-JAN-1999; 99US-0115142P.
XX PR 12-APR-1999; 99US-00290049.
XX PR 07-MAR-2002; 2002US-0363209P.
XX PR 08-AUG-2002; 2002US-0402483P.
XX PR 07-MAR-2003; 2003US-00383930.
XX PA (SMIT/) SMITH D J.
XX PA (TAUB/) TAUBMAN M A.
XX PI Smith DJ, Taubman WA;
XX PI WPI; 2005-151644/16.
XX DR
XX PT New composition comprising a fragment of a glucan binding protein-B
XX PT (GbpB) that binds to MHC class II protein, and a biocompatible
XX PT microparticle, useful for producing an antibody (claimed) for immunizing
XX PT mammals against dental caries.
XX PS Claim 7; SEQ ID NO 34; 73pp; English.
XX CC
XX CC The invention relates to a composition comprising a fragment of a glucan
XX CC binding protein-B (GbpB) and a biocompatible microparticle, where the
XX CC fragment binds to a major histocompatibility complex (MHC) class II
XX CC protein. The composition is useful for producing an antibody for
XX CC immunizing mammals against dental caries. This sequence corresponds to a
XX CC Streptococcus mutans GbpB protein of the invention.
XX SQ Sequence 1475 AA;
Query Match 100.0%; Score 7741; DB 9; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDKVKRYKLRKVKGRWTVSVASAVMTLTTLTSLGGLVKADSNESKQISNDSNTSVVTANE 60
DB 1 MDKVKRYKLRKVKGRWTVSVASAVMTLTTLTSLGGLVKADSNESKQISNDSNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPNYTVGETASNGEKLQNTT 120
DB 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPNYTVGETASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAQLILEKLPNVKEIDGKYVYDNNKGKVRT 180
DB 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAQLILEKLPNVKEIDGKYVYDNNKGKVRT 180
QY 181 NPTLADGKILHFEDETGYATDTSIDTVNKKDIDVTRSNLYKKYNQVYDRSAQSFVHDHYL 240
DB 181 NPTLADGKILHFEDETGYATDTSIDTVNKKDIDVTRSNLYKKYNQVYDRSAQSFVHDHYL 240

```

```

QY 241 TAESWYRPKYILKDGKTWTQSTEKDFRPLLTWPDQETQRYQYVNMQAQLGINKTYDDT 300
DB 241 TAESWYRPKYILKDGKTWTQSTEKDFRPLLTWPDQETQRYQYVNMQAQLGINKTYDDT 300
QY 301 SNQLQLNTAAATIQAIEBAKITTLKNTDMLRQTISAFVKTSAMNSDSEKPPDDHLQNGA 360
DB 301 SNQLQLNTAAATIQAIEBAKITTLKNTDMLRQTISAFVKTSAMNSDSEKPPDDHLQNGA 360
QY 361 VLYDNEGKLTYPYANSNYRIILNRTPTNQTGKDPRTYADNTTGGYEFLLANDVDNSNPVQ 420
DB 361 VLYDNEGKLTYPYANSNYRIILNRTPTNQTGKDPRTYADNTTGGYEFLLANDVDNSNPVQ 420
QY 421 AEQLNLHLFLNFGNIYANDPDANFDSIRVDADVNDADLLQIAGDYLKAAAGIHKNDKA 480
DB 421 AEQLNLHLFLNFGNIYANDPDANFDSIRVDADVNDADLLQIAGDYLKAAAGIHKNDKA 480
QY 481 ANDHLSILEAWSNDNTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMNPILTNSLV 540
DB 481 ANDHLSILEAWSNDNTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQRSGMNPILTNSLV 540
QY 541 NRTDNDASTAAVPSYFIRAHDSVQDLIADIIKAEINPNVVGYSFTMEEIKKAEIYNK 600
DB 541 NRTDNDASTAAVPSYFIRAHDSVQDLIADIIKAEINPNVVGYSFTMEEIKKAEIYNK 600
QY 601 DLLATEKKYTHYNTALSYALLLTNKSVPVRYGDMFTDDGOYMAHKTIINYEAIETLLKA 660
DB 601 DLLATEKKYTHYNTALSYALLLTNKSVPVRYGDMFTDDGOYMAHKTIINYEAIETLLKA 660
QY 661 RIKYVSGGOAMRNQGVNSEIITSVRYGKGALKATDGTDRTRTSGVAVIEGNNPSLRUK 720
DB 661 RIKYVSGGOAMRNQGVNSEIITSVRYGKGALKATDGTDRTRTSGVAVIEGNNPSLRUK 720
QY 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSQBAAGLVRYTNDRGELIFTAADIKG 780
DB 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSQBAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPOVSGVLGVVVPVGAALIKMPALRLARPHQOQWASVHONALDSRVNMFEGSFOAPA 840
DB 781 YANPOVSGVLGVVVPVGAALIKMPALRLARPHQOQWASVHONALDSRVNMFEGSFOAPA 840
QY 841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVSVSTSGSFLDSVILQNGYAFTRDYDLGIS 900
DB 841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVSVSTSGSFLDSVILQNGYAFTRDYDLGIS 900
QY 901 KPNKGTADDLVKAIKALHSKGIKVMADWPDQMTAFPEKEVVTATRDVKYGTVPAGSOI 960
DB 901 KPNKGTADDLVKAIKALHSKGIKVMADWPDQMTAFPEKEVVTATRDVKYGTVPAGSOI 960
QY 961 KNTLVVVDKSGKQQAQYKGAFLLEELQAKYPELFARKQISTGVPMDPSPVKIKOWSAKY 1020
DB 961 KNTLVVVDKSGKQQAQYKGAFLLEELQAKYPELFARKQISTGVPMDPSPVKIKOWSAKY 1020
QY 1021 FNGTNILGRGAGYVLKQATNTYFNI SDNKEINFLPKTLNLDQSDQVGSYDGKGVYVYST 1080
DB 1021 FNGTNILGRGAGYVLKQATNTYFNI SDNKEINFLPKTLNLDQSDQVGSYDGKGVYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYIFDNNYNGVMTGAQISNGVNYVYFLSNGLOLRDAILKNEDGTAY 1140
DB 1081 SGYQAKNTFISEGDKWYIFDNNYNGVMTGAQISNGVNYVYFLSNGLOLRDAILKNEDGTAY 1140
QY 1141 YCNDGRRYENGYYQFMSGVRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTTADCK 1200
DB 1141 YCNDGRRYENGYYQFMSGVRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTTADCK 1200
QY 1201 TRYFDKQSGNMYRNFRIENEGKWLILGEDGAATVTSQTINGQHLYFRANGVQVKGFEVY 1260
DB 1201 TRYFDKQSGNMYRNFRIENEGKWLILGEDGAATVTSQTINGQHLYFRANGVQVKGFEVY 1260
QY 1261 DHHGRISYDNGSGDQIRNRFVRNAQGWFFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
DB 1261 DHHGRISYDNGSGDQIRNRFVRNAQGWFFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
QY 1321 GBFVTVDRYGRISYDNGSGDQIRNRFVRNAQGWFFDNNGYAVTGARTINGQHLYFRAN 1380

```

Db 1321 GEFVTDYGRISYDNGSDQIRNRFVNAQGWYFDNNGYAVTGARTINGQHLYFRAN 1380
QY 1381 GVQVKGFEFTDRHGRISYDNGSDQIRNRFVNAQGWYFDNNGYAVTGARTINGQHLYFRAN 1440
Db 1381 GVQVKGFEFTDRHGRISYDNGSDQIRNRFVNAQGWYFDNNGYAVTGARTINGQHLYFRAN 1440
QY 1441 YFRANQVKGFEFTDRYGRISYDNGSDQIRNRFVNAQGWYFDNNGYAVTGARTINGQHLYFRAN 1475
Db 1441 YFRANQVKGFEFTDRYGRISYDNGSDQIRNRFVNAQGWYFDNNGYAVTGARTINGQHLYFRAN 1475

RESULT 4

AAU98030
ID AAU98030 standard; protein; 1475 AA.
XX
AC AAU98030;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFB mutant I448V.
XX
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; muten.
XX
OS Streptococcus mutans.
OS Synthetic.

PH Key Location/Qualifiers
FT Misc-difference 448 /note= "wild-type Ile substituted by Val"
FT US2002031826-A1.
XX
PN 14-MAR-2002.
XX
PD 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
PA (NICH/) NICHOLS S E.
XX
PI Nichols SE;
XX
WPI; 2002-414332/44.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
PS Claim 36; Page; 44pp; English.
XX

CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K790/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,

CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX

SQ Sequence 1475 AA;

Query Match 99.9%; Score 7740; DB 5; Length 1475;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1474; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDKVRVYKLRKVKRQWTVSVASAVMTLTTLGGVLKADSNESKQISNDSNTSVVTANE 60
Db 1 MDKVRVYKLRKVKRQWTVSVASAVMTLTTLGGVLKADSNESKQISNDSNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTSSSSTS VVNPKVSVNPTVGTETASNGEKLQNTT 120
Db 61 ESNVITEATSKQEAASSQTNHTVTSSSSTS VVNPKVSVNPTVGTETASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKOTTEADTDVDDSNAAANLQILEKLPNVKEIDGKYVYDNNKQVPT 180
Db 121 TVDKTSEAAANNISKOTTEADTDVDDSNAAANLQILEKLPNVKEIDGKYVYDNNKQVPT 180
QY 181 NFTLIADGKILHFDGTGAYTDTSIDTVNKDI VTTTRSNLYKKYQNVYDRAQSFHVDHYL 240
Db 181 NFTLIADGKILHFDGTGAYTDTSIDTVNKDI VTTTRSNLYKKYQNVYDRAQSFHVDHYL 240
QY 241 TAEVYRPRKYLKOGKWTQSTEDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT 300
Db 241 TAEVYRPRKYLKOGKWTQSTEDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT 300
QY 301 SNOLQNTAAATIQAKIEAKITTLKNTDMLRQTIISAFVKTQSAWNSDEKPPDDHLONGA 360
Db 301 SNOLQNTAAATIQAKIEAKITTLKNTDMLRQTIISAFVKTQSAWNSDEKPPDDHLONGA 360
QY 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTKKDPRTYADNTIGGYEFLANDVDSNPNVQ 420
Db 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTKKDPRTYADNTIGGYEFLANDVDSNPNVQ 420
QY 421 AEQLNWLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYKAAKGIHKNDKA 480
Db 421 AEQLNWLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYKAAKGIHKNDKA 480
QY 481 ANDHLSILEANSNDNTPYLHDDGDNMINNDKRLSLFLSLAKPLNQRSGMPLTNSLV 540
Db 481 ANDHLSILEANSNDNTPYLHDDGDNMINNDKRLSLFLSLAKPLNQRSGMPLTNSLV 540
QY 541 NRTDDNAETAAVPSYFTRAHDSEVQDIIADIIKAEINPNVVGYSFTWEEIKKAEIYNK 600
Db 541 NRTDDNAETAAVPSYFTRAHDSEVQDIIADIIKAEINPNVVGYSFTWEEIKKAEIYNK 600
QY 601 DLLATEKKTHTYNTALSYALLITNKSVPVYVGMFTDDGQYMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKTHTYNTALSYALLITNKSVPVYVGMFTDDGQYMAHKTINYEAIETLLKA 660

Db 601 DLATEKKYTHYNTALSYALLTNKSSVPRVYVYGMFTDDGQYMAHKTINVEAIEITLLKA 660
Qy 661 RIKYVSGGAMRNQOVGNSEIITSVRYKGKALKATDGTDRTRTSGVAVIEGNPSLRK 720
Db 661 RIKYVSGGAMRNQOVGNSEIITSVRYKGKALKATDGTDRTRTSGVAVIEGNPSLRK 720
Qy 721 ASDRVVNVNGAAHKNQVRPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Db 721 ASDRVVNVNGAAHKNQVRPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Qy 781 YANPOVSGYLVGVVFGAALIKWFAIRLARLPHQOQMASVHQNAALDSRVMEFSGSFQAF 840
Db 781 YANPOVSGYLVGVVFGAALIKWFAIRLARLPHQOQMASVHQNAALDSRVMEFSGSFQAF 840
Qy 841 TKKEBYTNVIAKNVDKFAEWGVTDFEMAPQVVSSTGDSFLDSVTQNGVAFTRDRLDGLIS 900
Db 841 TKKEBYTNVIAKNVDKFAEWGVTDFEMAPQVVSSTGDSFLDSVTQNGVAFTRDRLDGLIS 900
Qy 901 KPNKYGTADDLVKAIKALHSKGIKYMADWVPDMYAFPEKEVVTATRVVDKYGTPVAGSOI 960
Db 901 KPNKYGTADDLVKAIKALHSKGIKYMADWVPDMYAFPEKEVVTATRVVDKYGTPVAGSOI 960
Qy 961 KNTLVVDGKSSGKQQAQYKYGAFLEELQAKYFELFARKQISTGVPMPSVKIKQWSAKY 1020
Db 961 KNTLVVDGKSSGKQQAQYKYGAFLEELQAKYFELFARKQISTGVPMPSVKIKQWSAKY 1020
Qy 1021 FNGTNILGRGAGYVLKQATNTYFNI SDNKEINFLPKTLNODSOVGSYDGKGVYVYST 1080
Db 1021 FNGTNILGRGAGYVLKQATNTYFNI SDNKEINFLPKTLNODSOVGSYDGKGVYVYST 1080
Qy 1081 SGYQAKNTFISGDKWYFDNNGYVMTGAQSNNGVYFLSNGQLRLDAILKNEGDGVAY 1140
Db 1081 SGYQAKNTFISGDKWYFDNNGYVMTGAQSNNGVYFLSNGQLRLDAILKNEGDGVAY 1140
Qy 1141 YGNDGRRYENGYYQFMSGWRHFNNGEMSVGLTVIDGVOYFDEMGYQAKGFVTTADGK 1200
Db 1141 YGNDGRRYENGYYQFMSGWRHFNNGEMSVGLTVIDGVOYFDEMGYQAKGFVTTADGK 1200
Qy 1201 TRYFDKQSGNMYRNFRIENEGKWLVLGEBDGAATVTSOTINGOHLYFRANGVQVKGFEV 1260
Db 1201 TRYFDKQSGNMYRNFRIENEGKWLVLGEBDGAATVTSOTINGOHLYFRANGVQVKGFEV 1260
Qy 1261 DHHGRIISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGOLLYFRANGVQVK 1320
Db 1261 DHHGRIISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGOLLYFRANGVQVK 1320
Qy 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGOHLYFRAN 1380
Db 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGOHLYFRAN 1380
Qy 1381 GVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGOHLY 1440
Db 1381 GVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGOHLY 1440
Qy 1441 YFRANGVQVKGFEVTDYGRISYDNGSERVRIN 1475
Db 1441 YFRANGVQVKGFEVTDYGRISYDNGSERVRIN 1475

RESULT 5

AAU98040

ID AAU98040 standard; protein; 1475 AA.

XX AC

XX AAU98040;

XX DT

XX 27-AUG-2002 (first entry)

XX DE

XX S. mutans glucosyltransferase GTFB mutant K779Q.

XX KW

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;

XX coating composition; glucan; starch; latex; thermoplastic molecule;

XX KW

XX amyloplast; vacuole; paper manufacture; mutant; mutein.

XX XX

OS Streptococcus mutans.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 779 /note= "Wild-type Lys substituted by Gln"
XX
XX US2002031826-A1.
XX PD 14-MAR-2002.
XX PF 19-DEC-2000; 2000US-00740274.
XX PR 07-JUN-1995; 95US-00478704.
XX PR 07-JUN-1995; 95US-00482711.
XX PR 07-JUN-1995; 95US-00485243.
XX PR 16-JAN-1998; 98US-00007999.
XX PR 16-JAN-1998; 98US-00008172.
XX PR 20-JAN-1998; 98US-00009620.
XX PR 21-DEC-1998; 98US-00210361.
XX PA (NICH/) NICHOLS S E.
XX PI Nichols SE;
XX WPI; 2002-414332/44.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
XX as substitutes for and additions to modified starch and latexes in paper
XX manufacture, comprises mutations in specific positions.
XX
XX Claim 36; Page; 44pp; English.
XX
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
XX B polypeptide having changes at position from 1448V, D457N, D567T,
XX K1014T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
XX I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
XX GTF D polypeptide having changes at positions from T589D, T589E, N471D,
XX N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
XX GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
XX complementary polynucleotide, a ribonucleic acid sequence encoding the
XX GTF mutant, an expression cassette comprising the polynucleotide operably
XX linked to a promoter, a vector comprising the expression cassette, host
XX cell introduced with the vector, a transgenic plant comprising the
XX vector, a seed or tuber from the transgenic plant, a paper sizing and/or
XX coating composition comprising a glucan produced in a plant transformed
XX with a gene encoding the mutant GTF, wild type or, starch, a latex,
XX thermoplastic molecule or their combinations or glucan and starch where
XX the glucan is produced in the amyloplast and/or vacuole or a maize line
XX deficient in starch biosynthesis, transformed with a gene encoding a
XX glucosyltransferase B or D enzyme, wild-type or mutant and a paper
XX comprising the glucan (paper sizing/coating agent). The vector is useful
XX for producing a glucan in a plant. The method comprises transforming a
XX plant cell with the vector, growing the plant cell under plant growing
XX conditions to produce a regenerated plant and inducing expression of the
XX polynucleotide for a time sufficient to produce the glucan in the
XX regenerated plant, where the vector contains a transit sequence from
XX ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
XX chlorophyll AB binding protein to produce a transgenic plant, and glucan
XX is produced in the amyloplast of potato or the vacuole of sugar beet.
XX Glucans are useful as substitutes for and additions to modified starch
XX and latexes in paper manufacture. Unlike prior art techniques, which
XX require input materials that produce chemical effluents, paper
XX manufacture utilising the glucan produced by GTF, which utilises
XX biologically produced input materials, is more cost-effective and
XX environmentally friendly. Moreover, glucans also exhibit thermoplastic
XX properties and impart gloss to the paper during coating step. The present
XX sequence represents a GTFB mutant of the invention. Note: The present
XX sequence is not shown in the specification but was created by the indexer
XX using the GTFB sequence appearing as AAU98027 and the information in
XX claim 36
XX
XX Sequence 1475 AA;

Db 1381 GVOVKGEFVDRHGRISYYDGNSSGDIQIRNRFVRNAGQWFYFDNNGYAVTGARTINGOHL 1440

Qy 1441 YFRANGVQKGEFVDRYGRISYYDANSGERVRIN 1475
 |||||

Db 1441 YFRANGVQKGEFVDRYGRISYYDANSGERVRIN 1475
 |||||

RESULT 7
 AAU98033
 ID AAU98033 standard; protein; 1475 AA.
 XX AAU98033;
 XX
 XX 27-AUG-2002 (first entry)
 XX
 XX S. mutans glucosyltransferase GTFB mutant K1014T.
 XX
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 XX Streptococcus mutans.
 OS Synthetic.
 OS
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 1014
 FT /note= "Wild-type Lys substituted by Thr"
 XX
 XX US2002031826-A1.
 XX
 XX 14-MAR-2002.
 XX
 XX 19-DEC-2000; 2000US-00740274.
 XX
 XX 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 XX (NICH//) NICHOLS S E.
 PA
 XX
 XX Nichols SE;
 PI
 XX
 XX WPI; 2002-41432/44.
 DR
 XX
 XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 PT
 XX
 XX Claim 36; Page; 44pp; English.
 PS
 XX
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 1448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful

CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 XX
 XX Sequence 1475 AA;

Query Match 99.9%; Score 7735; DB 5; Length 1475;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1474; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MDKVRVYKLRVKRQWTVSVASAVMTLTLLSGGLVKADSNESKQISNDSTSVVTANE 60
 |||||

Db 1 MDKVRVYKLRVKRQWTVSVASAVMTLTLLSGGLVKADSNESKQISNDSTSVVTANE 60
 |||||

Qy 61 ESNVITEATSKQEAASSQTNHTVTTSSTSVYVNPKEVVSVPYTVGTASNGEKLQNTT 120
 |||||

Db 61 ESNVITEATSKQEAASSQTNHTVTTSSTSVYVNPKEVVSVPYTVGTASNGEKLQNTT 120
 |||||

Qy 121 TVDKTSEAAANNISKQTTTEADTDVDDSNAAANLQLEKLPNKVEIDGKYYYDNNKQVPT 180
 |||||

Db 121 TVDKTSEAAANNISKQTTTEADTDVDDSNAAANLQLEKLPNKVEIDGKYYYDNNKQVPT 180
 |||||

Qy 181 NFTLIADGKILHFDGTGAYTDTSIDTVNKDIIVTRSNLYKKYNQVYDRSAQSFHVDHYL 240
 |||||

Db 181 NFTLIADGKILHFDGTGAYTDTSIDTVNKDIIVTRSNLYKKYNQVYDRSAQSFHVDHYL 240
 |||||

Qy 241 TAESWYRKYILKDGKTWTQSTEDKDFRLLMTWPDQETQRYVYVNMNAQLGINKTYDDT 300
 |||||

Db 241 TAESWYRKYILKDGKTWTQSTEDKDFRLLMTWPDQETQRYVYVNMNAQLGINKTYDDT 300
 |||||

Qy 301 SNQLQNTAAATIAQKIEAKITTLKNTDWLQRTISAFVKTQSAMNSDSEKPPDDHLQNGA 360
 |||||

Db 301 SNQLQNTAAATIAQKIEAKITTLKNTDWLQRTISAFVKTQSAMNSDSEKPPDDHLQNGA 360
 |||||

Qy 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTGKDPRTYADNTIGGYEFLANDVDNSNPVQ 420
 |||||

Db 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTGKDPRTYADNTIGGYEFLANDVDNSNPVQ 420
 |||||

Qy 421 AEQLNWLHFLMNFNIYANDPDANFDSIRVDANVDNADLLQIAGDYLKAAGIHKNDKA 480
 |||||

Db 421 AEQLNWLHFLMNFNIYANDPDANFDSIRVDANVDNADLLQIAGDYLKAAGIHKNDKA 480
 |||||

Qy 481 ANDHLSILEAVSDNDTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQSGMPLTNSLV 540
 |||||

Db 481 ANDHLSILEAVSDNDTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQSGMPLTNSLV 540
 |||||

Qy 541 NRTDDNAETAAPVPSYFIRAHDSVQDLIADIILKAEINPNVYVGYFTWEEIKKAEIYNK 600
 |||||

Db 541 NRTDDNAETAAPVPSYFIRAHDSVQDLIADIILKAEINPNVYVGYFTWEEIKKAEIYNK 600
 |||||

Qy 601 DLLATEKKTHTYNALSVALLLTNKSSVPRVYVGYDMFTDDGOYMAHKTIINYEAIETLLKA 660
 |||||

Db 601 DLLATEKKTHTYNALSVALLLTNKSSVPRVYVGYDMFTDDGOYMAHKTIINYEAIETLLKA 660
 |||||

Qy 661 RIKVSGGQAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTTSVAVIEGNNPSLRUK 720
 |||||

Db 661 RIKVSGGQAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTTSVAVIEGNNPSLRUK 720
 |||||

manufacture, comprises mutations in specific positions.

Claim 36; Page; 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from I448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T, I448V/D457N/D567T/D571K/K7790/K1014T, Y169A/Y170A/Y171A, and K7790 or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

Sequence 1475 AA:

Query Match 99.8%; Score 7729; DB 5; Length 1475;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1473; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

	Qy	1	MDKKVRYKLRKYKRWVTVSVASAVMTLTTLGGGLXADSNESKSQISNDSENTSVVTANE	60
	Dd	1	MDKKVRYKLRKYKRWVTVSVASAVMTLTTLGGGLXADSNESKSQISNDSENTSVVTANE	60
	Qy	61	ESNVITEATSKOBAASSQTNIHTVTTSSSTSVVNPKEVSNPNYTVGETASNGEKLQNQTT	120
	Dd	61	ESNVITEATSKOBAASSQTNIHTVTTSSSTSVVNPKEVSNPNYTVGETASNGEKLQNQTT	120
	Qy	121	TVDKTSEAAANNISKOTTEADTDVIDDSNAANLOILEKLPNVXEIDCKYYYDNNCKVRT	180
	Dd	121	TVDKTSEAAANNISKOTTEADTDVIDDSNAANLOILEKLPNVXEIDGKYYYDNNCKVRT	180
	Qy	181	NFTLADGKILHFDETGAYTDTSIDTVNKDIVTTTRSNLYKKYNQVDRSAQSFEHVHYL	240
	Dd	181	NFTLADGKILHFDETGAYTDTSIDTVNKDIVTTTRSNLYKKYNQVDRSAQSFEHVHYL	240
	Qy	241	TAESWYRPKYILLKGKWTQSTEKDFRPLLMTWWPDOEQTORQYVYMNALGINKTYDDT	300
	Dd	241	TAESWYRPKYILLKGKWTQSTEKDFRPLLMTWWPDOEQTORQYVYMNALGINKTYDDT	300
	Qy	301	SNOQLNTAAATIQAIEAKITTLTKNTDWLRQTTTSFVKVQTOSAWNSDSEKFPDDHLONGA	360
	Dd	301	SNOQLNTAAATIQAIEAKITTLTKNTDWLRQTTTSFVKVQTOSAWNSDSEKFPDDHLONGA	360

Qy	361	VLVDNEGLTPYANSNRYILNRPTNQTKGKDRYTDADNTIGGYEFLLANDVDSNPVVQ	420
Db	361	VLVDNEGLTPYANSNRYILNRPTNQTKGKDRYTDADNTIGGYEFLLANDVDSNPVVQ	420
Qy	421	AEOLNWLHFLMNFNGIYANDPDANFOSIRYDADVNDVADLLOJTAGDYLKAAGHKNDKA	480
Db	421	AEOLNWLHFLMNFNGIYANDPDANFOSIRYDADVNDVADLLOJTAGDYLKAAGHKNDKA	480
Qy	481	ANDHLSILEAWSNDTPYLHDDGDNIMDNKLRLLSFLAKPLNORSGNPLITNSLV	540
Db	481	ANDHLSILEAWSNDTPYLHDDGDNIMDNKLRLLSFLAKPLNORSGNPLITNSLV	540
Qy	541	NRTDDNAETAAPVSYSFIRAHDSEVODLIADIIKAEINPNVGVSYFTMBEIKKAFEIYNK	600
Db	541	NRTDDNAETAAPVSYSFIRAHDSEVODLIADIIKAEINPNVGVSYFTMBEIKKAFEIYNK	600
Qy	601	DLLEATEKKYTHYNTALSYALLLTNKSSVPRVYVYGDMPFTDDGOYMAHKTINYEAEITLLKA	660
Db	601	DLLEATEKKYTHYNTALSYALLLTNKSSVPRVYVYGDMPFTDDGOYMAHKTINYEAEITLLKA	660
Qy	661	RIKYVSGGOAMRNOOVGNSEBIIISVRYGKALKATDTGDRTRTTSYGVAVIEGNNPSLRLK	720
Db	661	RIKYVSGGOAMRNOOVGNSEBIIISVRYGKALKATDTGDRTRTTSYGVAVIEGNNPSLRLK	720
Qy	721	ASDRVVMVMGAHKNQAYREPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGGELIFTAADIKG	780
Db	721	ASDRVVMVMGAHKNQAYREPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGGELIFTAADIKG	780
Qy	781	YANPOVSGYLGWVPVPGAALIKNFALRLARPHOOMASVHQNAALDSRVMEGFSNFOAPA	840
Db	781	YANPOVSGYLGWVPVPGAALIKNFALRLARPHOOMASVHQNAALDSRVMEGFSNFOAPA	840
Qy	841	TKXEEYTNVVIKKNVDKFAEWGVTDFEMAPQYVSSYSTDGSLDSVIONGYAFTDRYDUGIS	900
Db	841	TKXEEYTNVVIKKNVDKFAEWGVTDFEMAPQYVSSYSTDGSLDSVIONGYAFTDRYDUGIS	900
Qy	901	KPNKYGTADDLVKAIKALHSKGIKWADWVPDQMYAFPEKEVVTTATRDVKYGPVAGSQI	960
Db	901	KPNKYGTADDLVKAIKALHSKGIKWADWVPDQMYAFPEKEVVTTATRDVKYGPVAGSQI	960
Qy	961	KNTLYVVDGKSSGKDOQAKYGGAPLEBELQAKYPELFARKOISTGVPMDPSVKIKQWSAKY	1020
Db	961	KNTLYVVDGKSSGKDOQAKYGGAPLEBELQAKYPELFARKOISTGVPMDPSVKIKQWSAKY	1020
Qy	1021	FNGNTIILGRGAGYVLKDQANTYTFNISDNKEINFLPKTLNQDSQVGSYDGGYVYVYST	1080
Db	1021	FNGNTIILGRGAGYVLKDQANTYTFNISDNKEINFLPKTLNQDSQVGSYDGGYVYVYST	1080
Qy	1081	SGYQAKNTFISEGDKWYTFDNNGYMTGAQSIGVNYVYFLSNGLQLRDAIILKNEDGTYAY	1140
Db	1081	SGYQAKNTFISEGDKWYTFDNNGYMTGAQSIGVNYVYFLSNGLQLRDAIILKNEDGTYAY	1140
Qy	1141	YGNDGRYENGYYQFMGSVWRHFNNGEMSVGLTVIDGOVQYFDEMGGYQAKGKFFVTADGK	1200
Db	1141	YGNDGRYENGYYQFMGSVWRHFNNGEMSVGLTVIDGOVQYFDEMGGYQAKGKFFVTADGK	1200
Qy	1201	IRYFDKOSGNMYRNRFTIENEGKWLVLGEDGAAVTGSQTLNGOHLXFRANGVOVKGFBVT	1260
Db	1201	IRYFDKOSGNMYRNRFTIENEGKWLVLGEDGAAVTGSQTLNGOHLXFRANGVOVKGFBVT	1260
Qy	1261	DHGRISYYDGNSGDOIERNRFRVNAOQOWFYFDNNGYAVTGARTINGQLLYFRANGVOVK	1320
Db	1261	DHGRISYYDGNSGDOIERNRFRVNAOQOWFYFDNNGYAVTGARTINGQLLYFRANGVOVK	1320
Qy	1321	GEFVTDRYGRISYYDGNSGDOIERNRFRVNAOQOWFYFDNNGYAVTGARTINGOHLXFRAN	1380
Db	1321	GEFVTDRYGRISYYDGNSGDOIERNRFRVNAOQOWFYFDNNGYAVTGARTINGOHLXFRAN	1380
Qy	1381	GVQVKGFBVTDHGRISYYDGNSGDOIERNRFRVNAOQOWFYFDNNGYAVTGARTINGOHL	1440
Db	1381	GVQVKGFBVTDHGRISYYDGNSGDOIERNRFRVNAOQOWFYFDNNGYAVTGARTINGOHL	1440
Qy	1441	YFRANGVOVKGFBVTDRYGRISYYDANSGBSERVRIN	1475

QY 721 ASDRVVNMGAHKQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Db |||||
721 ASDRVVNMGAHKQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPOVSGYLVVVPVGAALIKWFAIRLARPHQQMASVHQNAALDSRVNFEQFSNFQAF 840
Db |||||
781 YANPOVSGYLVVVPVGAALIKWFAIRLARPHQQMASVHQNAALDSRVNFEQFSNFQAF 840
QY 841 TKKEEYTNVIAKNVDKFAENGVTDFEMAPQVVSSTDGSFLDSVIONGYAFTRDRLGIS 900
Db |||||
841 TKKEEYTNVIAKNVDKFAENGVTDFEMAPQVVSSTDGSFLDSVIONGYAFTRDRLGIS 900
QY 901 KPNKGTGTTADDLKAIALKHSKGIKWADWVPDMTAFPEKEVVTATRVDKYGTPTVAGSQI 960
Db |||||
901 KPNKGTGTTADDLKAIALKHSKGIKWADWVPDMTAFPEKEVVTATRVDKYGTPTVAGSQI 960
QY 961 KNTLVVVDGKSSGKQQAQKYGAFLEELQAKYPELFARKQISTGVPMDBPSVKIKQWSAKY 1020
Db |||||
961 KNTLVVVDGKSSGKQQAQKYGAFLEELQAKYPELFARKQISTGVPMDBPSVKIKQWSAKY 1020
QY 1021 FNGTNILGRGAGYVLKQDQNTYFNISDNKEINFLPKTLLNQDSQVGFSGYGVYVYST 1080
Db |||||
1021 FNGTNILGRGAGYVLKQDQNTYFNISDNKEINFLPKTLLNQDSQVGFSGYGVYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYTFDNNGYVMTGAQISNGVNYVFLSNGLQRLDAILKNEBDGTAY 1140
Db |||||
1081 SGYQAKNTFISEGDKWYTFDNNGYVMTGAQISNGVNYVFLSNGLQRLDAILKNEBDGTAY 1140
QY 1141 YGNDGRRYENGYYQPMGSGWRHFNNGEMSVGLTVIDGQVQVFDENGYQAKGKFTVTADGK 1200
Db |||||
1141 YGNDGRRYENGYYQPMGSGWRHFNNGEMSVGLTVIDGQVQVFDENGYQAKGKFTVTADGK 1200
QY 1201 IRYFDKQSGNMYRNFRIENEGKWLILGEDGAAVTGSQTINGQHLVFRANGVQVKGFEVT 1260
Db |||||
1201 IRYFDKQSGNMYRNFRIENEGKWLILGEDGAAVTGSQTINGQHLVFRANGVQVKGFEVT 1260
QY 1261 DHHGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQLLYFRANGVOVK 1320
Db |||||
1261 DHHGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQLLYFRANGVOVK 1320
QY 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQHLVFRAN 1380
Db |||||
1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQHLVFRAN 1380
QY 1381 GVQVKGFEVTDHGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQHL 1440
Db |||||
1381 GVQVKGFEVTDHGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQHL 1440
QY 1441 YFRANGVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQHL 1475
Db |||||
1441 YFRANGVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFDNNGYAVTGARTINGQHL 1475

RESULT 11
AAU98036
XX ID AAU98036 standard; protein; 1475 AA.
XX AC AAU98036;
XX 27-AUG-2002 (first entry)
XX DE S. mutans glucosyltransferase GTFB mutant D567T/D571K.
XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mitein.
XX OS Streptococcus mutans.
OS Synthetic.
XX FH Key Location/Qualifiers
FT Misc-difference 567
FT /note= "Wild-type Asp substituted by Thr"

FT Misc-difference 571
FT /note= "Wild-type Asp substituted by Lys"
XX US2002031826-A1.
XX PD 14-MAR-2002.
XX PF 19-DEC-2000; 2000US-00740274.
XX PR 07-JUN-1995; 95US-00478704.
XX PR 07-JUN-1995; 95US-00482711.
XX PR 16-JAN-1998; 98US-000485243.
XX PR 16-JAN-1998; 98US-00007999.
XX PR 16-JAN-1998; 98US-00008172.
XX PR 20-JAN-1998; 98US-00009620.
XX PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
FT as substitutes for and additions to modified starch and latexes in paper
FT manufacture, comprises mutations in specific positions.
XX Claim 36; Page; 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes p1 or p2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX SQ Sequence 1475 AA;

Query Match 99.8%; Score 7727; DB 5; Length 1475;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
PS
XX Claim 36; Page; 44pp; English.
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX
SQ Sequence 1475 AA;

Query Match 99.7%; Score 7721; DB 5; Length 1475;
Best Local Similarity 99.8%; Fred. No. 0;
Matches 1472; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKKVRYKLRKVKRQWTVSVASAVMTLTTLSSGLVKADSNESKQISNDSNTSVVTANE 60
DB 1 MDKKVRYKLRKVKRQWTVSVASAVMTLTTLSSGLVKADSNESKQISNDSNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPYTVGTASNGEKLQNTT 120
DB 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPYTVGTASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKOTTEADTVDDSNANLQILEKLPNVKEIDGKYYYDNNKVRT 180
DB 121 TVDKTSEAAANNISKOTTEADTVDDSNANLQILEKLPNVKEIDGKYYYDNNKVRT 180
QY 181 NFTLIADGKILHFDSTGAYTDTSDTVNKDIDVTRSNLYKKYNQVYDRSAQSFHVHDYL 240
DB 181 NFTLIADGKILHFDSTGAYTDTSDTVNKDIDVTRSNLYKKYNQVYDRSAQSFHVHDYL 240
QY 241 TAEWYRPKYILKDGKWTQSTEKDFRPLMTWPDQETQRQYVNMNAQLGINKTYDDT 300
DB 241 TAEWYRPKYILKDGKWTQSTEKDFRPLMTWPDQETQRQYVNMNAQLGINKTYDDT 300
QY 301 SNQLQLNTAATIQAIEAKITTLKNTDNLQRTISAFVKTSQAWNSDSEKPPDDHLQNGA 360

DB 301 SNQLQLNTAATIQAIEAKITTLKNTDNLQRTISAFVKTSQAWNSDSEKPPDDHLQNGA 360
QY 361 VLYDNEGKLTTPYANSNYRILNRTPTNQTGKKDPRTYADNTIGGYEFLLANDVDSNPVQ 420
DB 361 VLYDNEGKLTTPYANSNYRILNRTPTNQTGKKDPRTYADNTIGGYEFLLANDVDSNPVQ 420
QY 421 AEQLNLHFLMNFNIYANDPDANPDSIRVDADVNDADLLQIAGDYLKAAKGIHNKDA 480
DB 421 AEQLNLHFLMNFNIYANDPDANPDSIRVDADVNDADLLQIAGDYLKAAKGIHNKDA 480
QY 481 ANDHLSILEASDNDTPYLHDDGDNMINDKRLSLFLSLAKPLNQRSGMNPILTNSLV 540
DB 481 ANDHLSILEASDNDTPYLHDDGDNMINDKRLSLFLSLAKPLNQRSGMNPILTNSLV 540
QY 541 NRTDNDNAETAAVPSYSFIRAHDSEVQDLIADIIKAEINPNVVVGSFTMEEIKKAFIYNK 600
DB 541 NRTDNDNAETAAVPSYSFIRAHDSEVQDLIADIIKAEINPNVVVGSFTMEEIKKAFIYNK 600
QY 601 DLLATEKKYTHYNTALSYALLLTNNKSSVPRVYVYGMFTDDGOYMAHKIINYBAIETLLKA 660
DB 601 DLLATEKKYTHYNTALSYALLLTNNKSSVPRVYVYGMFTDDGOYMAHKIINYBAIETLLKA 660
QY 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDGTGRTTSTSGVAVIEGNNPSLRUK 720
DB 661 RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDGTGRTTSTSGVAVIEGNNPSLRUK 720
QY 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
DB 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANQVSGYLGWVPVGAALIKMFALRLARPHQQWASVHQAALDSRVMFEGSFQAPA 840
DB 781 YANQVSGYLGWVPVGAALIKMFALRLARPHQQWASVHQAALDSRVMFEGSFQAPA 840
QY 841 TKKEEYTNVTKXNDKFAEWGVTDFEWAPOVVSSTGDSFLDSVTLQNGVAFTRDRLDGLIS 900
DB 841 TKKEEYTNVTKXNDKFAEWGVTDFEWAPOVVSSTGDSFLDSVTLQNGVAFTRDRLDGLIS 900
QY 901 KPNKYGTADDLVKAIKALHSXGIKVMADWPDQMYAFPEKEVVTATRVDKYGTVPVAGSQI 960
DB 901 KPNKYGTADDLVKAIKALHSXGIKVMADWPDQMYAFPEKEVVTATRVDKYGTVPVAGSQI 960
QY 961 KNTLYVVDGKSSGKQQAQKYGGAFLLELOAKYPELFARKQISTGVMPDPSVKIKOWSAKY 1020
DB 961 KNTLYVVDGKSSGKQQAQKYGGAFLLELOAKYPELFARKQISTGVMPDPSVKIKOWSAKY 1020
QY 1021 FNGTNILGRGAGYVVKDQATNTYFNISDNKEINFLPKTLNODSOVGSYDGGKGVYVYST 1080
DB 1021 FNGTNILGRGAGYVVKDQATNTYFNISDNKEINFLPKTLNODSOVGSYDGGKGVYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYTFDNNGYMVTGAQISNGINGVNYFLSNGLQRLDAILKNEGTAY 1140
DB 1081 SGYQAKNTFISEGDKWYTFDNNGYMVTGAQISNGINGVNYFLSNGLQRLDAILKNEGTAY 1140
QY 1141 YGNDGRRYENGYYQFMSCGWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFPVTADGK 1200
DB 1141 YGNDGRRYENGYYQFMSCGWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFPVTADGK 1200
QY 1201 IRYFDKQSGNNMYRNFIRENEEGKWLILGEDGAAVTSOTINGOHLYFRANGVQVKEFVT 1260
DB 1201 IRYFDKQSGNNMYRNFIRENEEGKWLILGEDGAAVTSOTINGOHLYFRANGVQVKEFVT 1260
QY 1261 DHGGRISYDNGSGDQIRNRFVRNAQOGWYFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
DB 1261 DHGGRISYDNGSGDQIRNRFVRNAQOGWYFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
QY 1321 GEFVTDYGRISYYDNGSGDQIRNRFVRNAQOGWYFDNNGYAVTGARTINGOHLYFRAN 1380
DB 1321 GEFVTDYGRISYYDNGSGDQIRNRFVRNAQOGWYFDNNGYAVTGARTINGOHLYFRAN 1380
QY 1381 GVQVKEFVTDYGRISYYDNGSGDQIRNRFVRNAQOGWYFDNNGYAVTGARTINGOHL 1440

Db 1381 GVQVKGFEVTDHGRISYYDNGSGDQIRNFVRNAQGWFFDNGYAVTGARTINGQHL 1440

QY 1441 YPRANGVQVKGFEVTDHGRISYYDNGSGDQIRNFVRNAQGWFFDNGYAVTGARTINGQHL 1475

Db 1441 YPRANGVQVKGFEVTDHGRISYYDNGSGDQIRNFVRNAQGWFFDNGYAVTGARTINGQHL 1475

RESULT 13

ID AAU98039

AC AAU98039 standard; protein; 1475 AA.

XX AAU98039;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant YYY169-171AAA.

DE Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX

XX

PH Key Location/Qualifiers

FT Misc-difference 169..171

FT /note= "Wild-type Tyr-Tyr substituted by Ala-Ala"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

DR

XX Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.

XX

PS Claim 36; Page; 44pp; English.

XX

CC The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from 1448V, D457N, D567T,

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,

CC 1448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from 1589D, T589E, N471D,

CC N471D/T589E, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked to a promoter, a vector comprising the expression cassette, host

CC cell introduced with the vector, a transgenic plant comprising the

CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or

CC coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GTF, wild type or, starch, a latex,

CC thermoplastic molecule or their combinations or glucan and starch where

CC the glucan is produced in the amyloplast and/or vacuole or a maize line

CC deficient in starch biosynthesis, transformed with a gene encoding a

CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

CC comprising the glucan (paper sizing/coating agent). The vector is useful

CC for producing a glucan in a plant. The method comprises transforming a

CC plant cell with the vector, growing the plant cell under plant growing

CC conditions to produce a regenerated plant and inducing expression of the

CC polynucleotide for a time sufficient to produce the glucan in the

CC regenerated plant, where the vector contains a transit sequence from

CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and

CC chlorophyll AB binding protein to produce a transgenic plant, and glucan

CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC Glucans are useful as substitutes for and additions to modified starch

CC and latexes in paper manufacture. Unlike prior art techniques, which

CC require input materials that produce chemical effluents, paper

CC manufacture utilising the glucan produced by GTF, which utilises

CC biologically produced input materials, is more cost-effective and

CC environmentally friendly. Moreover, glucans also exhibit thermoplastic

CC properties and impart gloss to the paper during coating step. The present

CC sequence represents a GTFB mutant of the invention. Note: The present

CC sequence is not shown in the specification but was created by the indexer

CC using the GTFB sequence appearing as AAU98027 and the information in

CC claim 36

XX

SQ Sequence 1475 AA;

Query Match 99.7%; Score 7714; DB 5; Length 1475;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1472; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDKVVRYKLRVKRWVTVSVASAVMTLTTLTSLGGLVKADSNESKSOISNDSNTSVVTANE 60

Db 1 MDKVVRYKLRVKRWVTVSVASAVMTLTTLTSLGGLVKADSNESKSOISNDSNTSVVTANE 60

QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPYTVGTETASNGEKLQNTT 120

Db 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSNPYTVGTETASNGEKLQNTT 120

QY 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAANLQILEKLPVKEIDGKYVYDNNQKVRT 180

Db 121 TVDKTSEAAANNISKQTTTADTDVDDSNAAANLQILEKLPVKEIDGKYVYDNNQKVRT 180

QY 181 NFTLADGKILHFDGTGAYTDSIDTVNKDIIVTTRSNLYKYNQVYDRAQSPHEVDHYL 240

Db 181 NFTLADGKILHFDGTGAYTDSIDTVNKDIIVTTRSNLYKYNQVYDRAQSPHEVDHYL 240

QY 241 TAESWYRPKYILKDGKTWTQSTEKDFRPLLMTWPDQETQRYVYNNMAQLGINKTYDDT 300

Db 241 TAESWYRPKYILKDGKTWTQSTEKDFRPLLMTWPDQETQRYVYNNMAQLGINKTYDDT 300

QY 301 SNQLQNTAAATIOAKIEAKITTLKNTDHLROTISAFVKTQSAWNSDSEKPFDDHLLQNGA 360

Db 301 SNQLQNTAAATIOAKIEAKITTLKNTDHLROTISAFVKTQSAWNSDSEKPFDDHLLQNGA 360

QY 361 VLYDNEGKLTPTVANSNYRILNKTPTNQTGKDPRTADNTIGGYEFLANDVDSNPVQ 420

Db 361 VLYDNEGKLTPTVANSNYRILNKTPTNQTGKDPRTADNTIGGYEFLANDVDSNPVQ 420

QY 421 AEQLNLWFLMNFNGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLAAGKGIHKNDKA 480

Db 421 AEQLNLWFLMNFNGNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLAAGKGIHKNDKA 480

QY 481 ANDHLSILEAMSDNDTPYLHDDGDNINNDKRLSLFLPSLAKPLNQSGMPLIITNSLV 540

Db 481 ANDHLSILEAMSDNDTPYLHDDGDNINNDKRLSLFLPSLAKPLNQSGMPLIITNSLV 540

QY 541 NRTDNDNAETAAVPSYFIRAHDSVQDLIADIKAIEINPNVVGYSFTMEIHKAFIYNK 600

Db 541 NRTDNDNAETAAVPSYFIRAHDSVQDLIADIKAIEINPNVVGYSFTMEIHKAFIYNK 600

QY 601 DLLATEKYYTHNTALSVALLLTNKSSVPRVYVYGMFTDDGQYMAHKTINYEAIETLLKA 660

Db 601 DLLATEKYYTHNTALSVALLLTNKSSVPRVYVYGMFTDDGQYMAHKTINYEAIETLLKA 660

QY 661 RIKYVSGGQAMRNQVGNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNNPSLRK 720

Db 661 RIKYVSGGQAMRNQVGNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNNPSLRK 720

CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
XX claim 36
SQ Sequence 1475 AA;

Query Match 99.6%; Score 7711; DB 5; Length 1475;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1469; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MDKVRKYLKRVKRVVTVSVASAVMTLTTLTSLGLVKADSNESKSOISNDSNTSVVTANE	60
Db	1	MDKVRKYLKRVKRVVTVSVASAVMTLTTLTSLGLVKADSNESKSOISNDSNTSVVTANE	60
Qy	61	ENSVITEATSKQEAASSQTNHTVTTSSTSSVTVNPKVVSNTYTVGETASNGEKLQNTT	120
Db	61	ENSVITEATSKQEAASSQTNHTVTTSSTSSVTVNPKVVSNTYTVGETASNGEKLQNTT	120
Qy	121	TVDKTSEAAANNISQTTTADTDVDDSNAAQLILEKLPNVEIDGKYKYDNNNGKVRT	180
Db	121	TVDKTSEAAANNISQTTTADTDVDDSNAAQLILEKLPNVEIDGKYKYDNNNGKVRT	180
Qy	181	NETLLADGKILHFDGTGATDTSIDTVNKDIIVTTRSNLYKKNQVYDRSAQSFHVDHVL	240
Db	181	NETLLADGKILHFDGTGATDTSIDTVNKDIIVTTRSNLYKKNQVYDRSAQSFHVDHVL	240
Qy	241	TAESWYRPKYILKDGKTWTQSTKDFRPLLMTWPDQETQROYVYVMAQLGINKTYDDT	300
Db	241	TAESWYRPKYILKDGKTWTQSTKDFRPLLMTWPDQETQROYVYVMAQLGINKTYDDT	300
Qy	301	SNQLQNTAAATIOAKIEAKITTLKNTDMLRTISAFVKTSAMNSDSEKPPDDHLQNGA	360
Db	301	SNQLQNTAAATIOAKIEAKITTLKNTDMLRTISAFVKTSAMNSDSEKPPDDHLQNGA	360
Qy	361	VLYDNEGKLTPTVANSYRILNTPNQTCKDPRYTADNTICGYEFLLANDVDSNPVQ	420
Db	361	VLYDNEGKLTPTVANSYRILNTPNQTCKDPRYTADNTICGYEFLLANDVDSNPVQ	420
Qy	421	AEQLNLHFLMNFNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLKAAKGIHNKDKA	480
Db	421	AEQLNLHFLMNFNIYANDPDANFDSIRVDADVNDVADLLQIAGDYLKAAKGIHNKDKA	480
Qy	481	ANDHLSILEAASNDTPYLHDDGDMNMDNKLRLSLFLSLAKPLNQRSGMPLTNSLV	540
Db	481	ANDHLSILEAASNDTPYLHDDGDMNMDNKLRLSLFLSLAKPLNQRSGMPLTNSLV	540
Qy	541	NRTDDNAETAAPVPSYFIRAHDSQVODLIADIKAEPNVVGVYFTWEEIKKAEFIYNK	600
Db	541	NRTDDNAETAAPVPSYFIRAHDSQVODLIADIKAEPNVVGVYFTWEEIKKAEFIYNK	600
Qy	601	DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGMFTDDGOYMAHKTINYEAIETLKA	660
Db	601	DLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGMFTDDGOYMAHKTINYEAIETLKA	660
Qy	661	RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTSGVAVIEGNNPSRLK	720
Db	661	RIKYVSGQAMRNQOVGNSEIITSVRYGKGALKATDTGDRTRTSGVAVIEGNNPSRLK	720
Qy	721	ASDRVVVNMGAHKQVAPRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG	780
Db	721	ASDRVVVNMGAHKQVAPRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG	780
Qy	781	YANPOVSGVLGVVPGVGAALIKMFALRLARPHQOQWASVHQAALDSRVNPFEGSFQAPA	840
Db	781	YANPOVSGVLGVVPGVGAALIKMFALRLARPHQOQWASVHQAALDSRVNPFEGSFQAPA	840
Qy	841	TKXEEYTNVIAKNVDKFAEAGVTDTFEMAPQVVSSTGDSFLDSVITQNGYAFTRDYLGIS	900
Db	841	TKXEEYTNVIAKNVDKFAEAGVTDTFEMAPQVVSSTGDSFLDSVITQNGYAFTRDYLGIS	900
Qy	901	KPNKYGTADDLVKAIKALHSGIKVMADWVPDMQYAFPEKEVVTATRVKDYGTPTVAGSQI	960
Db	901	KPNKYGTADDLVKAIKALHSGIKVMADWVPDMQYAFPEKEVVTATRVKDYGTPTVAGSQI	960

Qy	961	KNTLYVVDGSKSGKQQAQKYGGAFLEELQAKYPELPFARKQISTGVPMDSVKIKOWSAKY	1020
Db	961	KNTLYVVDGSKSGKQQAQKYGGAFLEELQAKYPELPFARKQISTGVPMDSVKIKOWSAKY	1020
Qy	1021	FNGTNILGRGAGYVLKQQAATNTYFNISDNKEINFLPKTLNLDQSDQVGSYDGGKYVYYS	1080
Db	1021	FNGTNILGRGAGYVLKQQAATNTYFNISDNKEINFLPKTLNLDQSDQVGSYDGGKYVYYS	1080
Qy	1081	SGYQAKNTFISGDKWYFDNNGYVMTGAQSTNGVNYVFLSNGLOLRDAILKNEDEGTAY	1140
Db	1081	SGYQAKNTFISGDKWYFDNNGYVMTGAQSTNGVNYVFLSNGLOLRDAILKNEDEGTAY	1140
Qy	1141	YNGDGRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGQVQYFDEMGGYQAKGKFTVTADGK	1200
Db	1141	YNGDGRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGQVQYFDEMGGYQAKGKFTVTADGK	1200
Qy	1201	TRYFDKQSGNMYRNFRIENEGKWLVLGEDGAAVTSQTINGOHLVFRANGVQVKGFEVT	1260
Db	1201	TRYFDKQSGNMYRNFRIENEGKWLVLGEDGAAVTSQTINGOHLVFRANGVQVKGFEVT	1260
Qy	1261	DHGHGRISYDGNSGDOIIRNRFVRNAQGWYFEDNNGYAVTGARTINGOLLYFRANGVOVK	1320
Db	1261	DHGHGRISYDGNSGDOIIRNRFVRNAQGWYFEDNNGYAVTGARTINGOLLYFRANGVOVK	1320
Qy	1321	GEFVTDYGRISYDGNSGDOIIRNRFVRNAQGWYFEDNNGYAVTGARTINGOHLVFRAN	1380
Db	1321	GEFVTDYGRISYDGNSGDOIIRNRFVRNAQGWYFEDNNGYAVTGARTINGOHLVFRAN	1380
Qy	1381	GVQVKGFEVTDHGHGRISYDGNSGDOIIRNRFVRNAQGWYFEDNNGYAVTGARTINGOHL	1440
Db	1381	GVQVKGFEVTDHGHGRISYDGNSGDOIIRNRFVRNAQGWYFEDNNGYAVTGARTINGOHL	1440
Qy	1441	YFRANGVQVKGFEVTDYGRISYDGNSGDOIIRNRFVRNAQGWYFEDNNGYAVTGARTINGOHL	1475
Db	1441	YFRANGVQVKGFEVTDYGRISYDGNSGDOIIRNRFVRNAQGWYFEDNNGYAVTGARTINGOHL	1475

RESULT 15

AAU79284	AAU79284 standard; protein; 1476 AA.
ID	AAU79284
XX	AAU79284;
AC	AAU79284;
DT	13-AUG-2002 (first entry)
XX	Streptococcus mutans monoclonal antibody-related protein #1.
DE	Streptococcus mutans monoclonal antibody-related protein #1.
XX	Antibody; dental caries; water insoluble glucan synthetase; anti-caries;
KW	glucosyl transferase-B; immunotherapy.
OS	Streptococcus mutans.
PN	JP2002114709-A.
PD	16-APR-2002.
XX	04-OCT-2000; 2000JP-00304889.
XX	04-OCT-2000; 2000JP-00304889.
XX	(UYNI-) UNIV NIPPON.
XX	WPI; 2002-448885/48.
PT	Anti-carries agent composed of a monoclonal antibody against an inhibitory
PT	enzyme against water insoluble glucan synthetase of glucosyl transferase-
XX	B (GTF-B) of Streptococcus mutans.
PS	Claim 3; Page 13-16; 28pp; Japanese.
CC	The invention relates to a monoclonal antibody against dental caries and
CC	an anti-carries agent composed of a monoclonal antibody produced by

```
CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (PERM P-17566)
CC or mouse-hybridoma MHP136 (fPERM P-17567), against an enzyme having
CC inhibitive activity against water insoluble glucan synthetase of Glucosyl
CC transferase-B. The monoclonal antibody specifically inhibits water
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
CC transferase-B and is used in the immunotherapy of dental caries. This
CC sequence represents a Streptococcus mutans monoclonal antibody-related
CC protein
XX
SQ Sequence 1476 AA;
Query Match 98.5%; Score 7628.5; DB 5; Length 1476;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1459; Conservative 3; Mismatches 11; Indels 5; Gaps 2;
QY 1 MDKKVRYKLRKVKKRWTVSVASAVMTLTLSGGLVKADSNESKQISNDSNTSVVTANE 60
DB |
DB 1 MDKKVRYKLRKVKKRWTVSVASAVMTLTLSGGLVKADSNESKQISNDSNTSVVTANE 60
QY 61 ESNVITEATSKOEAAASSQTNHTVTSSSTSVVNPKEVVSNPYTVGETASNGEKLQNTT 120
DB |
DB 61 ESNVITEATSKOEAAASSQTNHTVTSSSTSVVNPKEVVSNPYTVGETASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKQTTSEADTDVDDSNAAANLQILEKLPNVKEIDGKYVYDNNKGVRT 180
DB |
DB 121 TVDKTSEAAANNISKQTTSEADTDVDDSNAAANLQILEKLPNVKEIDGKYVYDNNKGVRT 180
QY 181 NFTLIADGKILHFDGTGAYTDTSIDTVNKKDIIVTTRSNLYKKYNQVYDSQAOSFEHVDHYL 240
DB |
DB 181 NFTLIADGKILHFDGTGAYTDTSIDTVNKKDIIVTTRSNLYKKYNQVYDSQAOSFEHVDHYL 240
QY 241 TAESWYRPKYILKDGKWTQSTQTEKDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT 300
DB |
DB 241 TAESWYRPKYILKDGKWTQSTQTEKDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT 300
QY 301 SNQLQNLNAAATIAQIEAKITTLKNTDMLRQTIISAFVKTSAMNSDSEKPPDDLHONGA 360
DB |
DB 301 SNQLQNLNAAATIAQIEAKITTLKNTDMLRQTIISAFVKTSAMNSDSEKPPDDLHONGA 360
QY 361 VLYDNEGKLTVPANSNYRILNRTPTNQTKDPRYTADNTIGGYEFFLLANDVDNSNPVVQ 420
DB |
DB 361 VLYDNEGKLTVPANSNYRILNRTPTNQTKDPRYTADNTIGGYEFFLLANDVDNSNPVVQ 420
QY 421 AEQLNLWHLFNMFGNIYANDPANFDSIRVDADVNDVADLLQIAGDYILKAAGIHKNDKA 480
DB |
DB 421 AEQLNLWHLFNMFGNIYANDPANFDSIRVDADVNDVADLLQIAGDYILKAAGIHKNDKA 480
QY 481 ANDHLSILEANSNDNTPYLHDDGDNNINMDNKLRLSLFSLAKPLNQRSGMNPILTNSLV 540
DB |
DB 481 ANDHLSILEANSNDNTPYLHDDGDNNINMDNKLRLSLFSLAKPLNQRSGMNPILTNSLV 540
QY 541 NRTDQNAETAAPVPSYFIRAHDSVEQDILADIKAIEINPNVVGYSFTMEEIKKAPEIYNK 600
DB |
DB 541 NRTDQNAETAAPVPSYFIRAHDSVEQDILADIKAIEINPNVVGYSFTMEEIKKAPEIYNK 600
QY 601 DLLATEKKYTHYNTALSYALLTNKSSVPRVYVYVGMFTDDGQYMAHKTINYEAIETLLKA 660
DB |
DB 601 DLLATEKKYTHYNTALSYALLTNKSSVPRVYVYVGMFTDDGQYMAHKTINYEAIETLLKA 660
QY 661 RIKYVSGGQMRNQVGNSEIITSVRYGKALKATDTGDRTRTSGVAVIEGNPSLRILK 720
DB |
DB 661 RIKYVSGGQMRNQVGNSEIITSVRYGKALKATDTGDRTRTSGVAVIEGNPSLRILK 720
QY 721 ASDRVVNVNNGAAHKNQAYRPLLLTTDNGIKAVHSQEAAGLVRYTNDRGELIFTAADIKG 780
DB |
DB 721 ASDRVVNVNNGAAHKNQAYRPLLLTTDNGIKAVHSQEAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPOVSGYLGVVVPVGAALIKKFAIRLA---RPHQQWASVHQNAAALDSRVNFEFGSNFQ 837
DB |
DB 781 YANPOVSGYLGVVVPVGAADQ--DVRVAASTAPSTDGKSVHQNAAALDSRVNFEFGSNFQ 838
QY 838 AFATKKEEYTNVVIKKNVDKFAEWGVTDFEMAPQVVSSTGDSFLDSVLQNGYAFTRDYDL 897
|
```

```
DB 839 AFATKKEEYTNVVIKKNVDKFAEWGVTDFEMAPQVVSSTGDSFLDSVLQNGYAFTRDYDL 898
QY 898 GISKPNKYGTADDLVKAIKALHSHKGIKVMADWVPDQMYAFPEKEVVTATRVDKYGTTPVAG 957
DB |
DB 899 GISKPNKYGTADDLVKAIKALHSHKGIKVMADWVPDQMYAFPEKEVVTATRVDKYGTTPVAG 958
QY 958 SQIKNTLYVVDGKSSGKQQAQKYGGAFLEELQAKYPFELFARKQISTGYPMDSVVKIKQWS 1017
DB |
DB 959 SQIKNTLYVVDGKSSGKQQAQKYGGAFLEELQAKYPFELFARKQISTGYPMDSVVKIKQWS 1018
QY 1018 AKYFNGTNIILGRGAGYVLKDOATNTYFNI SNKKEINFLPKTLNODSOVGSYDCKGYVY 1077
DB |
DB 1019 AKYFNGTNIILGRGAGYVLKDOATNTYFNI SNKKEINFLPKTLNODSOVGSYDCKGYVY 1078
QY 1078 YSTSGYQAKNTFIISGDKWYVYFDNNGYVMTGAQSIINGVNYVYFSLNGLQLRDAILKNEBGT 1137
DB |
DB 1079 YSTSGYQAKNTFIISGDKWYVYFDNNGYVMTGAQSIINGVNYVYFSLNGLQLRDAILKNEBGT 1138
QY 1138 YAYYNDGRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGVOYFDEMGYQAKGKFPVTTA 1197
DB |
DB 1139 YAYYNDGRRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGVOYFDEMGYQAKGKFPVTTA 1198
QY 1198 DGKIRYFDKQSGNMYRNRFFIENEGKWLILGEDGAATGSGOTINGQHLYFRANGVOVKGE 1257
DB |
DB 1199 DGKIRYFDKQSGNMYRNRFFIENEGKWLILGEDGAATGSGOTINGQHLYFRANGVOVKGE 1258
QY 1258 FVTDHGHGRISYYDGNSGDQIRNRFVRNAQOGWFYFDNNGYAVTGARTINGQLLYFRANGV 1317
DB |
DB 1259 FVTDHGHGRISYYDGNSGDQIRNRFVRNAQOGWFYFDNNGYAVTGARTINGQLLYFRANGV 1318
QY 1318 QVKGEFVTDYGRISYIDGNSGDQIRNRFVRNAQOGWFYFDNNGYAVTGARTINGQHLYF 1377
DB |
DB 1319 QVKGEFVTDYGRISYIDGNSGDQIRNRFVRNAQOGWFYFDNNGYAVTGARTINGQHLYF 1378
QY 1378 RANGVOVKGEFVTDHGHGRISYYDGNSGDQIRNRFVRNAQOGWFYFDNNGYAVTGARTING 1437
DB |
DB 1379 RANGVOVKGEFVTDHGHGRISYYDGNSGDQIRNRFVRNAQOGWFYFDNNGYAVTGARTING 1438
QY 1438 QHLYFRANGVOVKGEFVTDYGRISYIDGNSGERVRIN 1475
DB |
DB 1439 QHLYFRANGVOVKGEFVTDYGRISYIDGNSGERVRIN 1476
```

Search completed: February 11, 2006, 19:14:37
Job time : 153.817 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 32.2941 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-34

Perfect score: 7741

Sequence: 1 MDKKVRYKLRVKKRWTVS.....DRYGRISYYDANSGERVRIN 1475

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5 COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/6 COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/H COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/PCTUS COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/RE COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7741	100.0	1475	2	US-09-007-999-2
2	7741	100.0	1475	2	US-09-210-361-2
3	7741	100.0	1475	2	US-09-740-274-2
4	5285	68.3	1375	2	US-09-210-361-4
5	5285	68.3	1375	2	US-09-740-274-4
6	3817.5	49.3	1430	2	US-09-008-172-2
7	3817.5	49.3	1430	2	US-09-210-361-6
8	3817.5	49.3	1430	2	US-09-740-274-6
9	3506	45.3	1577	1	US-08-793-824-2
10	2537	32.8	2057	2	US-09-499-203-2
11	2276	29.4	1781	2	US-09-604-957-3
12	2276	29.4	1781	2	US-09-995-749A-2
13	1685	21.8	545	2	US-09-604-957-4
14	1657	21.4	545	2	US-09-995-749A-10
15	1552	20.0	523	2	US-09-604-957-5
16	1547.5	20.0	522	2	US-09-995-749A-11
17	1379	17.8	535	2	US-09-604-957-7
18	1379	17.8	535	2	US-09-995-749A-13
19	1291.5	16.7	584	2	US-09-604-957-6
20	1289.5	16.7	584	2	US-09-995-749A-12
21	817	10.6	349	2	US-09-009-620-2
22	456	5.9	2710	1	US-08-480-604A-6
23	456	5.9	2710	1	US-08-405-496A-6
24	456	5.9	2710	2	US-08-915-136-6
25	456	5.9	2710	2	US-08-957-310-6
26	456	5.9	2710	2	US-10-011-366-6
27	456	5.9	2710	2	US-09-084-517-6

28	402	5.2	811	1	US-08-480-604A-7	Sequence 7, Appli
29	402	5.2	811	1	US-08-405-496A-7	Sequence 7, Appli
30	402	5.2	811	2	US-08-915-136-7	Sequence 7, Appli
31	402	5.2	811	2	US-08-957-310-7	Sequence 7, Appli
32	402	5.2	811	2	US-10-011-366-7	Sequence 7, Appli
33	402	5.2	811	2	US-09-084-517-7	Sequence 7, Appli
34	402	5.2	812	1	US-08-480-604A-29	Sequence 29, Appl
35	402	5.2	812	2	US-08-915-136-29	Sequence 29, Appl
36	402	5.2	812	2	US-09-084-517-29	Sequence 29, Appl
37	400	5.2	866	2	US-09-545-773-2	Sequence 2, Appli
38	400	5.2	866	2	US-10-222-038-2	Sequence 2, Appli
39	381	4.9	2366	1	US-08-480-604A-10	Sequence 10, Appl
40	381	4.9	2366	1	US-08-405-496A-10	Sequence 10, Appl
41	381	4.9	2366	2	US-08-915-136-10	Sequence 10, Appl
42	381	4.9	2366	2	US-08-957-310-10	Sequence 10, Appl
43	381	4.9	2366	2	US-10-011-366-10	Sequence 10, Appl
44	381	4.9	2366	2	US-09-084-517-10	Sequence 10, Appl
45	369.5	4.8	1231	2	US-08-714-741-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match	100.0%	Score 7741;	DB 2;	Length 1475;
Best Local Similarity	100.0%	Pred. No. 0;	Mismatches 0;	Indels 0;
Matches 1475;	Conservative 0;			Gaps 0;
QY	1	MDKKVRYKLRVKKRWTVSVASAVMTLTTLSSGLVKADSNESKQISNDSNTSVVTANE	60	
Db	1	MDKKVRYKLRVKKRWTVSVASAVMTLTTLSSGLVKADSNESKQISNDSNTSVVTANE	60	
QY	61	ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSVNPYTVGTASNGEKLQNTT	120	
Db	61	ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSVNPYTVGTASNGEKLQNTT	120	
QY	121	TVDKTSEAAANNISKQTTTEADTVDDSDNAANLQILEKLPVKEIDGKYYVYDNNKQVET	180	
Db	121	TVDKTSEAAANNISKQTTTEADTVDDSDNAANLQILEKLPVKEIDGKYYVYDNNKQVET	180	
QY	181	NFTLIADGKILHFDGTGAYTDTSDTVNKDITVTRSNLYKKYNQVYDRSAQSFHVDHYL	240	
Db	181	NFTLIADGKILHFDGTGAYTDTSDTVNKDITVTRSNLYKKYNQVYDRSAQSFHVDHYL	240	
QY	241	TAESWYRPKYILKDGKVTQSTQTEKDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT	300	
Db	241	TAESWYRPKYILKDGKVTQSTQTEKDFRPLMTWPDQETQRYVNMNAQLGINKTYDDT	300	
QY	301	SNQLQNLNAAATIAQIAEKITTLKNTDHLROTISAFVKTSQAWNSDEKPPDDHLQNGA	360	
Db	301	SNQLQNLNAAATIAQIAEKITTLKNTDHLROTISAFVKTSQAWNSDEKPPDDHLQNGA	360	
QY	361	VLYDNEGKLTVPYANSRYILNRTPTNQTKKDPRTADVTIGGYEFLLANDVDSNPVQV	420	

Db 361 VLYDNEGKLTVPYANSNYRLNTPNTQKKDPRYTADNTIGYEFLLANDVDSNPVQ 420
QY 421 AEQLNLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKA 480
Db 421 AEQLNLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKA 480
QY 481 ANDHLSILEAWSDNDTPYLDHDDGNMNDKRLSLFLSLAKPLNQRSGMNPPLTNSLV 540
Db 481 ANDHLSILEAWSDNDTPYLDHDDGNMNDKRLSLFLSLAKPLNQRSGMNPPLTNSLV 540
QY 541 NRTDNDNAETAAPVPSYFIRAHDSVQDLDIADIIKAEINPNVVGYSFTMEEIKKAEIYNK 600
Db 541 NRTDNDNAETAAPVPSYFIRAHDSVQDLDIADIIKAEINPNVVGYSFTMEEIKKAEIYNK 600
QY 601 DLLATEKKYTHYNTALSALLTNKSSVPRVYVYGMFTDDGOYMAHKTIINYEAIETLLKA 660
Db 601 DLLATEKKYTHYNTALSALLTNKSSVPRVYVYGMFTDDGOYMAHKTIINYEAIETLLKA 660
QY 661 RIKYVSGGOAMRNOQVGNSEIITSVRYGKGALKATDGTDRTRTSGVAVIEGNNPSLRK 720
Db 661 RIKYVSGGOAMRNOQVGNSEIITSVRYGKGALKATDGTDRTRTSGVAVIEGNNPSLRK 720
QY 721 ASDRVVNVNGAAHKNQAYRPLLLTTDNGIKAYHSDOEAGLVRYTNDRGELIFTAADIKG 780
Db 721 ASDRVVNVNGAAHKNQAYRPLLLTTDNGIKAYHSDOEAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPOVSGYLVGVVPGAALIKWFAIRLARPHQOMASVHQNAALDSRVNFEFNSFOAFA 840
Db 781 YANPOVSGYLVGVVPGAALIKWFAIRLARPHQOMASVHQNAALDSRVNFEFNSFOAFA 840
QY 841 TKKEBYTNVIAKNVDKFAEWGVTDPEMAPQVVSSTDSGSLDSVTONGVAFTRDRLDGLIS 900
Db 841 TKKEBYTNVIAKNVDKFAEWGVTDPEMAPQVVSSTDSGSLDSVTONGVAFTRDRLDGLIS 900
QY 901 KPNKGTADDLVKALKAHLSKGIKVMADVPDMYAFPEKEVVTATRVKDTPTVAGSQI 960
Db 901 KPNKGTADDLVKALKAHLSKGIKVMADVPDMYAFPEKEVVTATRVKDTPTVAGSQI 960
QY 961 KNTLVVVDGSKGQOQAKYGAPLEELQAKYPFELFARKQISTGVPMPPSVKIKOWSAKY 1020
Db 961 KNTLVVVDGSKGQOQAKYGAPLEELQAKYPFELFARKQISTGVPMPPSVKIKOWSAKY 1020
QY 1021 FNGTNILGRGAGYVLKQATNTYFNI SDNKEINFLPKTLNODSQVGSYDGKGYYVYST 1080
Db 1021 FNGTNILGRGAGYVLKQATNTYFNI SDNKEINFLPKTLNODSQVGSYDGKGYYVYST 1080
QY 1081 SGYQAKNTFISGDKWYFDNNGYVMTGAQSLNGVNYVFLSNGQLRDAILKNEPDGTAY 1140
Db 1081 SGYQAKNTFISGDKWYFDNNGYVMTGAQSLNGVNYVFLSNGQLRDAILKNEPDGTAY 1140
QY 1141 YGNDGRRYENGYYQPMGSRWHRFNGNEMSVGLTVIDGOVQYFDEMGYQAKGKFTVTTADGK 1200
Db 1141 YGNDGRRYENGYYQPMGSRWHRFNGNEMSVGLTVIDGOVQYFDEMGYQAKGKFTVTTADGK 1200
QY 1201 TRYFDKQSGNMYRNRFIENEGKWLVLGEDGAATVGSOTINGQHLYFRANGVQVKGFEVFT 1260
Db 1201 TRYFDKQSGNMYRNRFIENEGKWLVLGEDGAATVGSOTINGQHLYFRANGVQVKGFEVFT 1260
QY 1261 DHGHRISYDNGSGDQIRNRFVRNAQGWYFEDNNGYAVTGARTINGQLLYFRANGVQV 1320
Db 1261 DHGHRISYDNGSGDQIRNRFVRNAQGWYFEDNNGYAVTGARTINGQLLYFRANGVQV 1320
QY 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWYFEDNNGYAVTGARTINGQHLYFRAN 1380
Db 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWYFEDNNGYAVTGARTINGQHLYFRAN 1380
QY 1381 GVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFEDNNGYAVTGARTINGQHL 1440
Db 1381 GVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWYFEDNNGYAVTGARTINGQHL 1440
QY 1441 YFRANGVQVKGFEVTDYGRISYDANSGERVRIN 1475
Db 1441 YFRANGVQVKGFEVTDYGRISYDANSGERVRIN 1475

Db 1441 YFRANGVQVKGFEVTDYGRISYDANSGERVRIN 1475
RESULT 2
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No.: 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
Query Match 100.0%; Score 7741; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDKVVYKLRVKRWRVTVSVASAVMTLTTLSGGLVKADSNESQISDNTSNTSVVTANE 60
Db 1 MDKVVYKLRVKRWRVTVSVASAVMTLTTLSGGLVKADSNESQISDNTSNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVNPKFVSNPVTYVGTASNGEKLQNTT 120
Db 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVNPKFVSNPVTYVGTASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKQTEADTDVIDDSNAANLQILEKLPNVKEIDGKYVYDNNKGVRT 180
Db 121 TVDKTSEAAANNISKQTEADTDVIDDSNAANLQILEKLPNVKEIDGKYVYDNNKGVRT 180
QY 181 NFTLIADGKILHFDGTAYTDTSIDTVNKDITVTRSNLYKKYNQVYDRSAQSFHVHLYL 240
Db 181 NFTLIADGKILHFDGTAYTDTSIDTVNKDITVTRSNLYKKYNQVYDRSAQSFHVHLYL 240
QY 241 TAESWYRKYILKDGKWTQSTQTEKDFRLMTWPDQETQROYVNMNAQLGINKTYDDT 300
Db 241 TAESWYRKYILKDGKWTQSTQTEKDFRLMTWPDQETQROYVNMNAQLGINKTYDDT 300
QY 301 SNQLQNLAAATIOAKIEAKITTLKNTDWTQRTISAFVKTOSAMNSDSEKPPDDHLONGA 360
Db 301 SNQLQNLAAATIOAKIEAKITTLKNTDWTQRTISAFVKTOSAMNSDSEKPPDDHLONGA 360
QY 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTGKKDPRYTADNTIGGYEFLLANDVDSNPVQ 420
Db 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTGKKDPRYTADNTIGGYEFLLANDVDSNPVQ 420
QY 421 AEQLNLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKA 480
Db 421 AEQLNLHFLMNFNGNIYANDPDANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKA 480
QY 481 ANDHLSILEAWSDNDTPYLDHDDGNMNDKRLSLFLSLAKPLNQRSGMNPPLTNSLV 540
Db 481 ANDHLSILEAWSDNDTPYLDHDDGNMNDKRLSLFLSLAKPLNQRSGMNPPLTNSLV 540

```
QY 541 NRTDDNAETAAPSYSFIRAHDSVQDLIADIKAIEINPNVVGYSFTMBEIKKAFIYNK 600
Db 541 NRTDDNAETAAPSYSFIRAHDSVQDLIADIKAIEINPNVVGYSFTMBEIKKAFIYNK 600
QY 601 DLLATEKKYTHYNTALSYALLTNKSSVPRVYGDFTDDGQYMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKYTHYNTALSYALLTNKSSVPRVYGDFTDDGQYMAHKTINYEAIETLLKA 660
QY 661 RIKYVSGGQAMRNOQVGNSEIITSVRYGKALKATDTGDRTRTSGVAVIEGNNPSLRUK 720
Db 661 RIKYVSGGQAMRNOQVGNSEIITSVRYGKALKATDTGDRTRTSGVAVIEGNNPSLRUK 720
QY 721 ASDRVVNNGAHKNQAYRPLLTDTNGIKAYHSDQEAAGLVRYTNRDRELIFTAADIKG 780
Db 721 ASDRVVNNGAHKNQAYRPLLTDTNGIKAYHSDQEAAGLVRYTNRDRELIFTAADIKG 780
QY 781 YANPQVSGYLVGVVPGVGAALIKMFALRLARPHQOMASVHQNAALDSRVNPEGFSNFQAF 840
Db 781 YANPQVSGYLVGVVPGVGAALIKMFALRLARPHQOMASVHQNAALDSRVNPEGFSNFQAF 840
QY 841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVYSSTDGSFLDSVIONGYAFTDRYDLGTS 900
Db 841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVYSSTDGSFLDSVIONGYAFTDRYDLGTS 900
QY 901 KPNKYGTADDLVKAIKALHSKGIKNWADWPQMYAFPEKEVVTATRVDKYGTTPVAGSOI 960
Db 901 KPNKYGTADDLVKAIKALHSKGIKNWADWPQMYAFPEKEVVTATRVDKYGTTPVAGSOI 960
QY 961 KNTLYVVDGSKGQDOQAKYGAFLEELQAKYPELPARKQISTGVPMDSVKIKOWSAKY 1020
Db 961 KNTLYVVDGSKGQDOQAKYGAFLEELQAKYPELPARKQISTGVPMDSVKIKOWSAKY 1020
QY 1021 FNGTNLGRGAGVLLKQDQATNTYFNISDNKEINFLPKTLNODSQVGFSDGKGVVYST 1080
Db 1021 FNGTNLGRGAGVLLKQDQATNTYFNISDNKEINFLPKTLNODSQVGFSDGKGVVYST 1080
QY 1081 SCYQAKNTFISEGDKWYIFDNNGYVMVTGAQSGINGVNYFSLNGLQLRDAILKNEGTAY 1140
Db 1081 SCYQAKNTFISEGDKWYIFDNNGYVMVTGAQSGINGVNYFSLNGLQLRDAILKNEGTAY 1140
QY 1141 YGNDGRRYENGYYQFMSGWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGFVTTADGK 1200
Db 1141 YGNDGRRYENGYYQFMSGWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGFVTTADGK 1200
QY 1201 IRYFDKQSNMYNRNFIENEKGWLYLGEDGAATVGSQTINGHLYFRANGVQVKEFVT 1260
Db 1201 IRYFDKQSNMYNRNFIENEKGWLYLGEDGAATVGSQTINGHLYFRANGVQVKEFVT 1260
QY 1261 DHHGRI SYDGNSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQLLYFRANGVQVK 1320
Db 1261 DHHGRI SYDGNSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQLLYFRANGVQVK 1320
QY 1321 GEFVTDYGRISYYDGNSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQHLYFRAN 1380
Db 1321 GEFVTDYGRISYYDGNSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQHLYFRAN 1380
QY 1381 GVQVGEFVTDYGRISYYDGNSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQHL 1440
Db 1381 GVQVGEFVTDYGRISYYDGNSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQHL 1440
QY 1441 YFRANGVQVKEFVTDYGRISYYDANSGERVRIN 1475
Db 1441 YFRANGVQVKEFVTDYGRISYYDANSGERVRIN 1475
```

RESULT 3

US-09-740-274-2

; Sequence 2, Application US/09740274

; Patent No. 6465203

; GENERAL INFORMATION:

; APPLICANT: Nichol, Scott E.

; TITLE OF INVENTION: Glucan-containing Compositions and Paper

```
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-740-274-2
```

```
Query Match 100.0%; Score 7741; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDKVKRYKLRKVKRWVTVSVASAVMTLTLSGLVKADSNESKSIQSNDSNTSVVTANE 60
Db 1 MDKVKRYKLRKVKRWVTVSVASAVMTLTLSGLVKADSNESKSIQSNDSNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVNPKVSVNPTVGETASNGEKLQNTT 120
Db 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVNPKVSVNPTVGETASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISQTTTEADTVLDDSNAAANLQILEKLPNVEIDGKYYYDNNNGKVRT 180
Db 121 TVDKTSEAAANNISQTTTEADTVLDDSNAAANLQILEKLPNVEIDGKYYYDNNNGKVRT 180
QY 181 NPTLIADGKILHFDGTGAYTDTSIDTVANKDI VTTTSNLYKKNQVYVDRSAQSFHVDHYL 240
Db 181 NPTLIADGKILHFDGTGAYTDTSIDTVANKDI VTTTSNLYKKNQVYVDRSAQSFHVDHYL 240
QY 241 TAESWYRKYILKDKGTWTQSTEDKFRPLMTWDPQETQRYVYVNMNAQLGINKTYDDT 300
Db 241 TAESWYRKYILKDKGTWTQSTEDKFRPLMTWDPQETQRYVYVNMNAQLGINKTYDDT 300
QY 301 SNQLQNLTAATIQAKIEAKITTLKNTDNLROTISAFVKTSQAMNSDSKPFDDHLQNGA 360
Db 301 SNQLQNLTAATIQAKIEAKITTLKNTDNLROTISAFVKTSQAMNSDSKPFDDHLQNGA 360
QY 361 VLYDNEGKLTVPYANSNYRILNRTPTNQTGKDPRTADNTIGGYEFLLANDVNSNPVQ 420
Db 361 VLYDNEGKLTVPYANSNYRILNRTPTNQTGKDPRTADNTIGGYEFLLANDVNSNPVQ 420
QY 421 AEQLNLHFLNFGNIYANDPDANPDSIRVDAVDNVDADLLQIAGDYKAAKGIIHKNDKA 480
Db 421 AEQLNLHFLNFGNIYANDPDANPDSIRVDAVDNVDADLLQIAGDYKAAKGIIHKNDKA 480
QY 481 ANDHLSIILEAWSDNDTPYLHDDGDNWINDNKLRLSLFLSLAKPLNQRSGMNPITNSLV 540
Db 481 ANDHLSIILEAWSDNDTPYLHDDGDNWINDNKLRLSLFLSLAKPLNQRSGMNPITNSLV 540
QY 541 NRTDDNAETAAPSYSFIRAHDSVQDLIADIKAIEINPNVVGYSFTMBEIKKAFIYNK 600
Db 541 NRTDDNAETAAPSYSFIRAHDSVQDLIADIKAIEINPNVVGYSFTMBEIKKAFIYNK 600
QY 601 DLLATEKKYTHYNTALSYALLTNKSSVPRVYGDFTDDGQYMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKYTHYNTALSYALLTNKSSVPRVYGDFTDDGQYMAHKTINYEAIETLLKA 660
```

```
Qy 661 RIKYVSGGQMRNQGVSEIITSVRYGKALKATDGTGDRTRTSGVAVIEGNNPSLRUK 720
Db 661 RIKYVSGGQMRNQGVSEIITSVRYGKALKATDGTGDRTRTSGVAVIEGNNPSLRUK 720
Qy 721 ASDRVVNNVGAAGKQVAPLPLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Db 721 ASDRVVNNVGAAGKQVAPLPLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Qy 781 YANPOVSGVLGVVVPVGAALIKQFALRLARPHQOQMASVHONAAALSRVMFEGFSFQAPA 840
Db 781 YANPOVSGVLGVVVPVGAALIKQFALRLARPHQOQMASVHONAAALSRVMFEGFSFQAPA 840
Qy 841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVVSSTGDSFLDSVLQNGVAFTRDYLIGS 900
Db 841 TKKEEYTNVIAKNVDKFAEWGVTDFEMAPQVVSSTGDSFLDSVLQNGVAFTRDYLIGS 900
Qy 901 KPNKYGTADDLKAIKALHSGIKVMADWVPDMYAFPEKEVYVTRVDKYGTPVAGSQI 960
Db 901 KPNKYGTADDLKAIKALHSGIKVMADWVPDMYAFPEKEVYVTRVDKYGTPVAGSQI 960
Qy 961 KNTLVVDGKSSGKQQAQYGGAFLEELQAKYPFARKQISTGVPMDPSPVKIKOWSAKY 1020
Db 961 KNTLVVDGKSSGKQQAQYGGAFLEELQAKYPFARKQISTGVPMDPSPVKIKOWSAKY 1020
Qy 1021 FNGTNILGRGAGYVLKDOATNTYFNI SDNKEINFLPKTLNODSQQVGSYDGKGYVYST 1080
Db 1021 FNGTNILGRGAGYVLKDOATNTYFNI SDNKEINFLPKTLNODSQQVGSYDGKGYVYST 1080
Qy 1081 SGYQAKNTFISEGDKWYIFDNNGVMTGAQSGINGVNYFLSNGLQLRDAILKNEGTYAY 1140
Db 1081 SGYQAKNTFISEGDKWYIFDNNGVMTGAQSGINGVNYFLSNGLQLRDAILKNEGTYAY 1140
Qy 1141 YNGDGRYENGYYQFMGQVWRHFNNGEMSVGLTVIDGQVQYVPEDEMGYQAKGKFVTTADGK 1200
Db 1141 YNGDGRYENGYYQFMGQVWRHFNNGEMSVGLTVIDGQVQYVPEDEMGYQAKGKFVTTADGK 1200
Qy 1201 TRYFDKQSNMYRNRFIENEKGWLYLGEDGAATVGTSGTQINGHLYFRANGVQVGFVT 1260
Db 1201 TRYFDKQSNMYRNRFIENEKGWLYLGEDGAATVGTSGTQINGHLYFRANGVQVGFVT 1260
Qy 1261 DHHGRISYDNGSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
Db 1261 DHHGRISYDNGSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQLLYFRANGVQVK 1320
Qy 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQHLYFRAN 1380
Db 1321 GEFVTDYGRISYDNGSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQHLYFRAN 1380
Qy 1381 GVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQHL 1440
Db 1381 GVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQHL 1440
Qy 1441 YFRANGVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQHL 1475
Db 1441 YFRANGVQVKGFEVTDYGRISYDNGSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQHL 1475
```

```
RESULT 4
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
```

```
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match 68.3%; Score 5285; DB 2; Length 1375;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1031; Conservative 101; Mismatches 169; Indels 59; Gaps 14;

Qy 1 MDKKVRYKLRKVKKRWVTVSVASAVMTLTLTSLGGLVKADSNESKSOISNDSNTSVVTANE 60
Db 1 MEKKVRFKLRKVKKRWVTVSVASAVMTLTLTSLGSLVKADSTDDRQQAFTESQASLVTTSE 60
Qy 61 -----ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSNPYTVGETA-- 109
Db 61 AAKETLTATDSTATSATSQPTA--TVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKT 118
Qy 110 SNGEKLQNTTVTDKTSSEA-----AANNISKQTTSEAD-----TDVIDDSN 149
Db 119 DNSDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKI VPIKPKIGKLKQPSLSQDDI 178
Qy 150 AANLQILKLPNKVEIDGKYVYDNNQKVRNFTLIAAGKILHPDETGAYTDTSDTVNK 209
Db 179 AA----LGNVKNIIRKVGKYYKXEDGTQKYNALNINGKTFFFDETGALSNTLPSKKG 234
Qy 210 DIVTT--RENLYKKNQVYDRSAQSPHEVDHYLTAEVSWRPKYILKDGKTWTOSTEKDPRP 268
Db 235 NITNDNTNSFAQYNQVYSTDVANPEHVDHYLTAEVSWRPKYILKDGKTWTOSTEKDPRP 294
Qy 269 LLMTWVPOEQTQRYVNTMNAQLGINKTYDDTSNQLQNLNIAAATOAKIEAKITTLKNTD 328
Db 295 LLMTWVPOEQTQRYVNTMNAQLGIHQYNTATSPQLNLAATOITQTKIEEKITAENNTN 354
Qy 329 WLQRTISAFVKTQSNWNSDSEKPPDDHLQNGAVLYDNEGKLTTPYANSNYRILNRPPTQOT 388
Db 355 WLQRTISAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQSAANSYRILNRPPTQOT 414
Qy 389 GKQDPRYTADNTIGGYEFLLANDVDNSNPVQAEQLNWLHFLMNFNGNIYANDPDANFDSI 448
Db 415 GKQDPRYTADNTIGGYEFLLANDVDNSNPVQAEQLNWLHFLMNFNGNIYANDPDANFDSI 474
Qy 449 RVDADVNDADLLOIAGDYLKAAGIKHNDKAANDHLSILEAWSNDPTPYLHDDGDNNMIN 508
Db 475 RVDADVNDADLLOIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDTPYLHDDGDNNMIN 534
Qy 509 MDNKLRLSLFLSLAKPLNORSQMPPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVDL 568
Db 535 MDNRLRLSLYSLAKPLNQRSQMPPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVDL 594
Qy 569 IADIIKAEINPNVVGYSFTMEIEIKKAPRIYKNDLLATEKKYTHYNTALSYALLTNKSSV 628
Db 595 IRNIIRTEINPNVVGYSFTTEIEIKKAFIYKNDLLATEKKYTHYNTALSYALLTNKSSV 654
Qy 629 PRVYVYDMDFTDGGQYMAHKTINYEAETLLKARIKYVSGGQAMRNOQVNSBIITSVRYG 688
Db 655 PRVYVYDMDFTDGGQYMAHKTINYEAETLLKARIKYVSGGQAMRNOQVNSBIITSVRYG 714
Qy 689 KGALKATDTGDRTRTSGVAVIEGNNPSLRKASDRVVVNNCAAHKNQAYRPLLTDTNG 748
Db 715 KGALKATDTGDRTRTSGVAVIEGNNPSLRKASDRVVVNNCAAHKNQAYRPLLTDTNG 774
Qy 749 IKAVHSDQEAAGLVRYTNDRGELIFTAADIKGANPOVSGYLVGVVVPVGAALIKQFALRL 808
```

Db 775 IKAYHSDQEAAGLVRVTNRDRELIFTAADIKGANPQVSGYLGWVPVGAADQ--DVRV 832
QY 809 A---RPHQQWASVHQNAALDSRVWFGFSNFQAFATKCEYTNVIAKNVDKFAEWGVT 865
Db 833 AASTASTDCKSVHQNAALDSRVWFGFSNFQAFATKCEYTNVIAKNVDKFAEWGVT 892
QY 866 FEMAPQVSTGDSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKV 925
Db 893 FEMAPQVSTGDSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKV 952
QY 926 MADWVPDQMYAPPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYGAFL 985
Db 953 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYGAFL 1012
QY 986 EELQAKYPELFARKQISTGVPMDPSVKIKQWSAKYFNGTNILGRGAGYVLKQATNTYFN 1045
Db 1013 EELQAKYPELFARKQISTGVPMDPSVKIKQWSAKYFNGTNILGRGAGYVLKQATNTYFN 1072
QY 1046 -TSDNKEINFLPKTLN-----QDSOVGFSYDGKGVYVYSTSGYQAKNTFISEGDKWYVF 1099
Db 1073 LVSDN---TFLPKSLVNPNGHTSSVTGLVFDGKGVYVYSTSGYQAKNAFISLGNWYVF 1129
QY 1100 DNNQYMTCAQSTNGWYVFLSNGQLRDAILKNGEDTYAYYCNDRRYENGYYQPMGCV 1159
Db 1130 DNNQYMTCAQSTNGWYVFLSNGQLRNAIYDNGKNVLSYTGNDGRRYENGYYLF-GQQ 1188
QY 1160 WRHFNNGEMSVGLTWDGQVFDGMYQAKGFVTADGKIRYFDPKQSGNMYRNFIE 1219
Db 1189 WYFQNGINAVGLTRVHGAVQYFDASGFOAKGQFITADGKIRYFDRDSGNQISNRFVN 1248
QY 1220 EBGKWLYLEGGAATGSGTQINGHLYFRANGVQKGFVTPDHGHSIYYDNGSGDQIRN 1279
Db 1249 SKGEWFLPHNGVAVGTGTENGQRLYFKPNGVQAKGEFIRDANGLYRYDPSNGNEVN 1308
QY 1280 RFRNACQGFYEDNNGYAVTGARTING-----QLLYPR 1313
Db 1309 RFRNRSKGWFLPHDNGIAVTGARVVVNGHASILSLMVFR 1347

RESULT 5
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CDD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 68.3%; Score 5285; DB 2; Length 1375;
Best Local Similarity 75.9%; Pred. No. 0;

Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;
QY 1 MDKQRYKLRKVKRWVTVSVASAVMTLTTLTSLGLVKADSNESKSIQISNDSTSVVTANE 60
Db 1 MEKQVRPKLRKVKRWVTVSVASAVMTLTTLTSLGSLVKADSTDDRQAVTESQASLVTTSE 60
QY 61 -----ESNVITEATSKOBAASQTNHTVTTSSTSVVNPKEWENPYTVGSETA-- 109
Db 61 AAKETLTATDTSTATSATSOPTA--TVTDNVSTTQSTNTTANTANFVVKPTTTSQAKT 118
QY 110 SNGEKLQNTTVDKTSSEA-----AANNISKOTTEAD-----TDVTDSDN 149
Db 119 DNSDKIITTSKAVNRLTATGKFPANNTAHPKTVTDKIVPIKPKIKGLKQPSLSQDDI 178
QY 150 AANLQILEKLPNKEIDGKYVYDNNKVRTNFTLIADCKILHFDSTGAVTDSIDTVNK 209
Db 179 AA-----LGNVKIRKNGKYYKYKEDGTLQKYNALNINKTFFDFTGALSNNLTSPSKG 234
QY 210 DIVTT-RGNLYKKYQVDRSQAQSPHEVDHYLTAEVWYRPKYILKDGKTWTOSTEKDRP 268
Db 235 NITNDNTNSPAQYNQVSTVDVANFEVDHYLTAEVWYRPKYILKDGKTWTOSTEKDRP 294
QY 269 LLMTWPDQETQRYVNTMNAQLGINKYVDDTSNOLNIAAATQAKIEAKITTLKNTD 328
Db 295 LLMTWPDQETQRYVNTMNAQLGINKYVDDTSNOLNIAAATQAKIEAKITTLKNTD 354
QY 329 WLQRTISAFVKTQSASWNSDEKPPDDHLQNGAVLYDNEGKLTPIYANSNRYILNRPTNQT 388
Db 355 WLQRTISAFVKTQSASWNSDEKPPDDHLQNGAVLYDNEGKLTPIYANSNRYILNRPTNQT 414
QY 389 GKQDPRTADNTIGGYEFLANDVNSNPVQAEQLNMLHFLMNFNIGNIYANDPDANFDSI 448
Db 415 GKQDPRTADNTIGGYEFLANDVNSNPVQAEQLNMLHFLMNFNIGNIYANDPDANFDSI 474
QY 449 RVDADVNDADLLQIAGDYLKAAGIHKNDKAANDHLSILEAWSNDTPYLHDDGDNMIN 508
Db 475 RVDADVNDADLLQIAGDYLKAAGIHKNDKAANDHLSILEAWSNDTPYLHDDGDNMIN 534
QY 509 MDNKLRLSLFLSLAKPLNORSQGNPLITNSLVNRTDDNAETAAPVSPYSFIRAHDSVQDL 568
Db 535 MDNKLRLSLFLSLAKPLNORSQGNPLITNSLVNRTDDNAETAAPVSPYSFIRAHDSVQDL 594
QY 569 IADIKAIEINPNVVGYSFTMBEIKKAFIYNKDLATEKKYTHYNTALSYALLNTKSSV 628
Db 595 IENIIRTEINPNVVGYSFTMBEIKKAFIYNKDLATEKKYTHYNTALSYALLNTKSSV 654
QY 629 PRVYVYDMDTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSBIITSRYG 688
Db 655 PRVYVYDMDTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSBIITSRYG 714
QY 689 KGALKATDTGDRTTSTGAVTEGNPISRLKASDRVVVNMAAHHQYRPLLLTTDNG 748
Db 715 KGALKATDTGDRTTSTGAVTEGNPISRLKASDRVVVNMAAHHQYRPLLLTTDNG 774
QY 749 IKAYHSDQEAAGLVRVTNRDRELIFTAADIKGANPQVSGYLGWVPVGAALIKMFAFL 808
Db 775 IKAYHSDQEAAGLVRVTNRDRELIFTAADIKGANPQVSGYLGWVPVGAADQ--DVRV 832
QY 809 A---RPHQQWASVHQNAALDSRVWFGFSNFQAFATKCEYTNVIAKNVDKFAEWGVT 865
Db 833 AASTASTDCKSVHQNAALDSRVWFGFSNFQAFATKCEYTNVIAKNVDKFAEWGVT 892
QY 866 FEMAPQVSTGDSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKV 925
Db 893 FEMAPQVSTGDSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKV 952
QY 926 MADWVPDQMYAPPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYGAFL 985
Db 953 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYGAFL 1012
QY 986 EELQAKYPELFARKQISTGVPMDPSVKIKQWSAKYFNGTNILGRGAGYVLKQATNTYFN 1045
Db 1013 EELQAKYPELFARKQISTGVPMDPSVKIKQWSAKYFNGTNILGRGAGYVLKQATNTYFN 1072

FILING DATE: 24-AUG-1994
TELECOMMUNICATION INFORMATION:
TELEPHONE: 61 2 9957 5944
TELEFAX: 61 2 957 6288
TELEX: 26547
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1577 amino acids
TYPE: amino acids
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Streptococcus salivarius

US-08-793-824-2
Query Match
Best Local Similarity 45.3%; Score 3506; DB 1; Length 1577;
Matches 733; Conservative 239; Mismatches 468; Indels 162; Gaps 28;
QY 1 MDKKVRYKLRKVKRWTVSVASAVMTLFTLSGGLV----KADSNEKS-----QISN 49
DB 1 MENKVRFKLHKVKONWTVIGVTT--LSMVALAGSSLLAQGVKVEADETSAPNGDGLQQLSE 58
QY 50 DNTSVVTANEESNVITEATSKQ-----EASQSTNHTVTTSSSTSV 92
DB 59 DGTASLVT-----TTTVTQASQAQSVASAVATASVSHTSFQAATSVAVSQEAQAQTSVP 114
QY 93 VNPKEVSNPYTVGTASNGEKLQ-----QTSQSGQETQTEQVSQQTSTQVAGQTSQSTPSVTSQAPRVL 169
DB 115 ASQEVAVSS-----QTSQSGQETQTEQVSQQTSTQVAGQTSQSTPSVTSQAPRVL 169
QY 117 -----NOTTVTDKTSEAAANNI-----SKOTTEAD 141
DB 170 NAAPAATRAADSTIRANRNWNTITASGTTPTNVTITGTPKPNVTVTSPNGTRPN 229
QY 142 TDVIDDSNAANLQILEKLP-----NVKEIDGKYVYDNNKGVRT 180
DB 230 VTIVTQPNKPNVPQSPQSPQPNQPNQSLDKPVASNLKTIQDKQYVE-NGVVK 288
QY 181 NPTLADCKILHFDGTGAVTDSIDTVNKDITVTESNLKYNQVYDRAQSFEHVDHL 240
DB 289 NAAIELDGLRYFDETGAVWQSKPLRADATIPNNS-IYAVYNQAYDTSKSSFEHLNPL 347
QY 241 TAESWYRPKYILKDGKTWTQSTEDKFRPLMTWPDQETQROQYVNYMAQLGINKTYDDT 300
DB 348 TADSWYRPKYILKDGKNWTASTEDKFRPLMTWPDQETQROQYVNYMAQLGINKTYDDT 407
QY 301 SNQLQNLTAATIAQIAEKITTLKNTDWRQTIISAFVKTQSAMNSDEKPP---DDHLQ 357
DB 408 MMSYDLAAAAETVQRIEERIGREGNTTTLRLQMSDFIKTQPMWSESEDLNLLVGKDLH 467
QY 358 NGAVLYDNEGKLTYPANSNYRLARTPTNOTKQKDPRTADNTICGYEFLLANDVDSNP 417
DB 468 GGAULTFLN-NSATSHANSDFRLMNRTPNTQGTGR--KYHIDRSNGGYELLANDINDSNP 524
QY 418 VVQAEQLNWLHPLMNFNGNIYANDPANFDSIRVDADVNDVADLLQIAGDYKKAAGIHK 477
DB 525 AVQAEQLNWLHYIMNIGILNDPNSANFGVDRIDAVNDVADLLQIASDYFKEKTRVADN 584
QY 478 DKAANDHLSILEAMSDNPTPYLHDDGDNMIMNDKRLSLFLSLAKPLNQRSGMPLITN 537
DB 585 EANAIAHLSILEAWSYNDHQNKQKGAQLSIDNPLRETLTTLFLRKSNGYRSLERVITN 644
QY 538 SLVNRDNDNAETAAPVPSFTRAHDSVQDLADIIEKAEINPVNVGVYSFTWEEIKAPEI 597
DB 645 SLNRRSSQKHPTPDANTIFVRAHDSVQAVLANIISQINPKTDGFTTMDLQKAEI 704
QY 598 YNKDLATEKKTHTYNTALSVALLLTNKSSVPRVYVYDGMFTDDGQYMAHKTINYEAITL 657
DB 705 YNADIADKADKTYQNIIPAAATMLTNKSDITRVYVYDGMFTDDGQYMAEKSPYNAIDAL 764
QY 658 LKARIKYVSGQAMRNQOVGNSEIITSVRYGKALKATDGTGTRTTRTSQVAVIEGNPNSL 717

DB 765 LRARIKYVAGQDMKVTKLNGVEIMSSVRYGKABEANQLGTAETRNQGMVLVLTANRPM 824
QY 718 RLKASDRVVVNNGAHKQOAYEPLLLTTDNGIKAYHSQDEA-AGLVRYTNDRGELIFTAA 776
DB 825 KUGANDRLVNNGAHKQOAYEPLLLSKTSKSTGLATYKSDSDVPAGLVRYTNDQNLFTTAD 884
QY 777 DIKGYANPOVSGYLVVVPVGAALIKMFALRLARPHQOQMASVHQNAALDSRVNFEFGSNF 836
DB 885 DIAGHSTVEVSGYLAHVVPVGAASENQDARTKASTKKGEQVFESSAALDSQVIYEGFSNF 944
QY 837 QAFATKKEBYTNVIAKQVDFKFAEKGVTDFEMAPQVVSSTDSGLDSVLTONGYAFTRDYD 896
DB 945 QDFVKTPSQYTNRVIAQNAKLFKENGITSFEFAPQVVSQDGTFLDSIIENGYAFEDRYD 1004
QY 897 LGISKNPKYGTADDLIVKAIKALHSKGIKVMADQVDPQVAFPEKEVVTATRDVYKGYTPVA 956
DB 1005 IAMSXNNKYGSLKDLMDALRALHAGISAIADVDPQIYNLPGKEVVTASRTNSYGTGPR 1064
QY 957 GSQIKNTLYVVDGKSSGDKQAKYGGAFLEELQAKYPELFAKQISTGVPMPSVKIKOW 1016
DB 1065 NAEIYNSLYAAKTRTFGNDFOGKYGGAFLDELKAKYPAIFERVQISNGRKLITNEKITOW 1124
QY 1017 SAKYFNGNIIKRGAGYVVKDQATNTYFNISDNKINFLPKTLTNQDSQVGSYDGKGYV 1076
DB 1125 SAKYFNGSNIQGTGARYVLQDNATNQYFSVKAQO--TFLPKQ-MTEITGSGFRVGGDDVQ 1181
QY 1077 YYSTSGYQAKNTFISEG-DKMYVFDNNGVMTGASINGVNYVFLSNGLQDLRALKNED 1135
DB 1182 YLSIGGILAKNTFIQVGANQMYVFDKNGMNMVTEQVVDGKKYFFFLDNGLQLRHVLQSGD 1241
QY 1136 GTYAYVNGDGRYENGYYQFMSSGVN--RHFN-NGEMSVGLTVIDGOVQYFDE-MGYQAKG 1191
DB 1242 GHVYVYDPKGVQAFNGFYDFAGPRQDVRYFDGNGQMYRGLHDMYGTTFYFDEKTIQAKD 1301
QY 1192 KPVTTADGKIYRFDKQSGNMYRNRIENEBGK-WLYLGEDGAATGSGQTINGOHLYFRAN 1250
DB 1302 KPIRPADGRTRYFIPDTGNLAVNRPAQNPKAMWYLDNSGVAVTGLQTINGKQYVFDNE 1361
QY 1251 GVQVKGFEVTDHGRISYYDNGSDQIRNRFVRNAQOGQVFDNNGYAVTGARTINGQLL 1310
DB 1362 GRQVKGHFEVTTNNQRY-FLDGDGSGEIAPSRPV-TENNKMYYVDGNGKLVKGAQVINGNHY 1419
QY 1311 YFRANGVQVKGFEVTDYGRISYYDNGSDQIRNRFVRNAQOGQVFDNNGYAVTGARTI 1370
DB 1420 YFNNDYSQVKGAWANGR-----YVDGDSQAVSNQFIQIAANQWAYLQDGHKVTGLQNI 1474
QY 1371 NGQHLVFRANGVQVKGFEVTDYGRISYYDNGSDQIRNRFVRNAQOGQVFDNNGYAVT 1430
DB 1475 NNKVVYFGSNGAQVKGKLLTVQ-GKKCYFDAHTGBOVVNRFEAAARGCWYYFNSAGQAVT 1533
QY 1431 GARTINGQHLVFRANGVQVKGFEVTDYGRISYYDNGSDQIRNRFVRNAQOGQVFDNNGYAVT 1470
DB 1534 QGVVINGKQLYFDGSGRQVKGRYV---YVGGKRLFCDAKTGE 1572

RESULT 10
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057

```

; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match      32.8%; Score 2537; DB 2; Length 2057;
Best Local Similarity 39.8%; Pred. No. 2.4e-161;
Matches 607; Conservative 205; Mismatches 486; Indels 228; Gaps 46;

Qy 93 VNPKEVSNPVYTGTSANGE-----KLQNTTTVDKTSSEAAANNISKOTTTEADTVIDDS 148
Db 286 VNNKRIYFAPNT-GNAVANTEIINGKLOGRDANGQVKNAPSKDVAGNTFFPDANGV--- 341
Qy 149 NAANLQILLEKLPVKEIDKYYVDNNKVRNFTPLIADGKLLHPD-ETGAYTDSIDTV 207
Db 342 -----MLTGLQTISGKTYLDEQGHLRKNYAGTFFNQPMYFDADTGA-GKTAIYBQ 391
Qy 208 NKDIVTTRSNLYKKYN--QVYDRSAQSPEHVDHYLTAESWYRKPVKILKDGKTKWTOSTEKD 265
Db 392 PDQGLVSQSNENTPHNAASKSYDKS--SPENVGGLYLTADTWYRPTDLKNGDWTASTETD 449
Qy 266 FRPLMTWMPQDETQORQVYNNYNAQ- LGINKTYDPTSQLOLNIAAATIQAKIEAKITTL 324
Db 450 MRPLMTWMPDKQTQANLYNFMSSKGLGITTYYTAATSKTLNDAAFVIQTAIEQOISLK 509
Qy 325 KNTDHLRQTIASFVKYTSQAWNSDSE-KPPD--DHLONGAVLYDNEGKLTTPYANS-NYRIL 380
Db 510 KSTELRLDAIDSFVKTOANWNNKQTEDEAFGLQWLQGGFLAYQDDSHRTPNPTDGNRRKL 569
Qy 381 NRTPTNQTGKDPRTYADNTTIGGYEFFLLANDVDNSNPVQAEQNLNHLFNMFGNIYAND 440
Db 570 GROPINIGSKD---TTGK--GSEFLLANDIDNSNPVQAEQNLNHLNFMFGSITGN 624
Qy 441 PDANPDSIRDAVONVDADLLQIAGDYLKAAGTHKNDKAANDHLSLEAASDNDTPYLH 500
Db 625 DNANPDIRDAVONVDADLLQIAGDYFKALYGTGDKSDANANKHLSILEEDWNGKDPQVY 684
Qy 501 DGDGMNIMNDKRLSLFLSLAKPLNQRSGM-----NPLITNSL 539
Db 685 QQGAQLTMDYTVTSQFQNSLTHGANNRSMWYFLDTGYLNGDLNKKIVDKNRPNSGTL 744
Qy 540 VNRTDDNAETAAPSPYSIRAHDSVQDLIADIKAIEINPNVVGY---SFTWEEIKKAFE 596
Db 745 VNRANSGETKVIYNYSFVRADHYDAQD---PIRKAMIDHGLIKNMQDTTFDQLAQOME 801
Qy 597 IYNKOLLATE--KKYTHYNTALSAYLLTNKSSVPRVYVYGMFTDDGQYMAHKTINYBAI 654
Db 802 FYKQENPSPGFKKYNDYNLPSAYAMLTKDTPRVYVYGMFTDGLBGGQYMEKGTIYNPVI 861
Qy 655 ETLLKARIKYVSGGQAM-----RNQOVNSEIITSRYGKALKATDTGDRTT----- 702
Db 862 SALLKARIKYVSGGQMTATDSSGDKLQGETDLSVRFGRGIM---TSQDTTQDQNSQ 917
Qy 703 --RTSGVAVIEGNPRLRKASDRVVVNMGAHKNQAVRPLLLTTDNGIKAYHSDQEAAG 760
Db 918 DYKQOGIIVGNPDLKNDKTTLHNGKAKHKNQLYRALVLSNDSGLDIVDSDDKAPT 977
Qy 761 LVRYTNRGELIFTAA-----DINGYANPQVSGYGVVVPVGAALIKMFALR 807
Db 978 L--RTNDNGDLIFHKNTNFVKQDGTI INYEMKGSINALISGLGVVVPVGAS----- 1027
Qy 808 LARPHQOQVASY-----HONAAALDSRVMEFGFSNFPQAPATKKEEYTNVVIK 854
Db 1028 ---DSQDARTVATESSSSSNDGVSFHSNAALDSNVIEGFSNFPQAMPPTSPEQSTNVVIATK 1084
Qy 855 VDKFAEWGTVDFEMAPQVSVSDG-----SFLDSVIQNGYAFTRDYDLIGISK-----PN 903
Db 1085 ANLFKELGITSELPAPQTRSSGDTYNGMSFLDSFLNNGYAFTRDYDLGFKADGNPPT 1144
Qy 904 KYGTADDLVKAIKALHSGKIKVMADVPDQMYAFPEKEVWVATRVKDYGTVPVAGSQIKXT 963
Db 1145 KYGTDDDLRNAIEALHKGMOAIDWVPDQIYALPGKEVWVATRVKDYGTVPVAGSQIKXT 1204
Qy 964 LYVVDGKSSGKDOQAQYGAFLLELQAKYPELFARKQISTGVMPDPSVKIKQWSAKYFNG 1023
```

```

Db 1205 LYVANTKSSGVYQAKYGGFEFLDKLREYPSLFQKNQVSTQPIDASTKIQWSAKYFNG 1264
Qy 1024 TNLGRGAGYVLKQOATNTYFNISDNKEINFLPKTLANQDSQVGSFYDGKGVYVYSTGY 1083
Db 1265 TNLHRCGAYYVLKQWATNWOYENIAKTNEV-FLPLQLQNKDAQTGTFISDASGVKYYSISGY 1323
Qy 1084 QAKTFTFISEG-KWYFFDNNGYMTGAQSGINGV-----NYFLSNGLOLRDAI 1130
Db 1324 QAKDTFIEDNGNMYFDKOGYMYRSQQGPNPIRTVETSVNTRNGNYFFMPNGVELRKGF 1383
Qy 1131 LKNEDGTVAYVGNDRRYENGYQFMGVMWRHEN-NGEMSVGLTVIDGQ-VQYFDEMVGQ 1188
Db 1384 GTDSNGVYFFDDQGMVRDKIINDDANNFVHLNVDGMSRGLFKFSDTLOYPASNGVQ 1443
Qy 1189 AKGFVTTADKIRYFDKQSGNMYNRNRFIENEEGK---W-----LYLGDGAAVTQSOTIN 1241
Db 1444 IKDSYAKDSKGNKYFDSATGN-----NDTGKATWDGNGYVITIDSDA-NNTIGVN 1494
Qy 1242 GQHLVFRANGVQVGEFVTDHHRISYYDNGSGD-----QIRNRFV-- 1282
Db 1495 TDYTAITYTSSUREDLFANAPYGVVT-KDQNGNDLKWOYINHTKQYEGQOVQVTRQYDTS 1553
Qy 1283 -----RNAQGOVEYEDNNGYAVTGARTINGOLLYF---RANGVQVKGEFVTD 1326
Db 1554 KGVSNLITFAGGDLQGORLWDSRALTWTPTFKTMN-QISFISYANRNDGLFLNAPYQVK 1612
Qy 1327 RYGRISYYDNGSGQIRNRFVRNAQGO-WFYFDNNG---YAVTGARTINGQH---LYFR 1378
Db 1613 GYQLAGHSNQYKGOQVTIAGVANVSGKDWLSLSFNGTQYWDISQALNTNFTDMNQKVFV 1672
Qy 1379 ANGVOVGEFVTDHHRISYYDNGSGDOIRNRFVRNAQGOVYFDNNG---YAVTGART 1434
Db 1673 NTTSLNDGLFLNAPYRPGYKLAGLAKYNNQTVTVSQQ---YFDDQGTVMSSQVVLGGQT 1729
Qy 1435 I-----NGOHLVFRANG 1446
Db 1730 VVVDNHALLAQMVSDTDOQLYVNSNG 1755

RESULT 11
US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

Query Match      29.4%; Score 2276; DB 2; Length 1278;
Best Local Similarity 42.4%; Pred. No. 4.2e-144;
Matches 518; Conservative 183; Mismatches 382; Indels 140; Gaps 37;

Qy 35 LVKADSNESKQIQSNDSTSVVVTANESNVITEATSKOEAAASSQTN--HTVTTSSSTSV 92
Db 134 IVLVNGKEVRQLVND-----TKGAAAGFNRNDYVKNPAINSSSM 174
Qy 93 VNPKEVSNPVY-GETASNGEKLNQNTT-----VDKTSBAAAANNISKOTTTEADTVID 146
```

Db 175 SGFGIITLPVTVKENVOLVHRFSNDVKTGEGNVDFWSEL-----MPVKD 221

Qy 147 DSNAAQLILEKLPNVKEIDGKYIYD--NNGKVRNFTLIADGKILHFDGTGAYTDTSD 205

Db 222 SFQKNGPL--KQFGLQTINGQOYIIDPTTGPQRKNFLQSGNNWIYFD-----SDTGVG 274

Qy 206 TVNKDIV-----TTRSN-LYKKNQVYDRSAQSFHVHDHYLTAEBSWYRPKYILKDGKTWQ 260

Db 275 TNALEQFAKGTVSSNEQYRNGNAAYSDDKSIENVNGYLTADTWYRPQKILKDGTTWD 334

Qy 261 STEKDFRPLMTWPDQETORQYVYNNAAQLGINKT-----YDDTSNQLNATAATQAK 316

Db 335 SKETDMRPLMVWPNLTLTQAYLYNMKQHGMLLPALPFFNADADPAELNHYSEIVQON 394

Qy 317 IEAKITTLKNTDMLRTQISAFVKTQSAMNSDSEK-PFDD-HLQNGAVLYDNEGKLTTPYAN 374

Db 395 IEKRISSETGNTDMLRTLMHDFVTNNPMKNKDSENVNFSGIQQGGFLKYEN-SDLTTPYAN 453

Qy 375 SNYRILNRTPNTQTKKDPRYTADNTIGGYEFLLANDVNSNPVQAEOLNHLFLMNFQ 434

Db 454 SDYRLLRGMPIN-----IKQDTYRQGEFLLANDIDNSNPVQAEOLNHLFLMNFQ 504

Qy 435 NIYANDPDANFDSIRVDADVNDADLLQIAGDYLAAGKIHKNDAANDHLSILEAWSN 494

Db 505 TITANNDQANFDSVRVDAPDNIDADLNNIAQDYFNAAAYGM-DSDAVSNKHINILEDWNHA 563

Qy 495 DTPYLLHDDGNNMINMDNKLRLSLFLSLAKPLNORSQGNPLITNSLVNRTDDNAETAAPVS 554

Db 564 DPEYFNKIGNPOLTWDDTIKNSLNHGLSDATN-RWGLDAIHQSADRENNSTENVVIFN 622

Qy 555 YSFIKRAHSEVDQLIADIKAIEINPNVVG---YSFTMBEIKKAFIYNKDLATEKKYTH 611

Db 623 YSFIKRAHSDNSQDQONAIR-----DVTGKYHTFTFEDEQKIDAYIQDQNSTVKKYNL 677

Qy 612 YNTALSYALLTNKSSVRVYGDFTDDGQYMAKNTINYEAETILLKARIKYVSGQAM 671

Db 678 YNIPASYALLTNKDTIPRVYGDYTDGQYMEHQTRYDYLTLNLLSRVKYKVGQSGM 737

Qy 672 RNOQV-GENEITTSVRYGKALKATDGTDRTRTSGVAVIEGNPFLSLKASDRVVVWNG 730

Db 738 QTMSSVGGNNILTSVRYGKAMATDGTGDETRTQGIQVVSNTFNLKLVNDKVLHNG 797

Qy 731 AAHKNQAYRPLLTDTNGIKAYHSDQEAAGLVRYTNDRGELIFTA-----AD--I 778

Db 798 AAHKNQAYRAAVLTDTTGDVINTSDQAGP--VAMTDENGDLVLSHNLVWNGKEADTAV 855

Qy 779 KYANPOVSYGLVWVPVGAALIKMFAURLA---RPHQOMASVHONAAALDSRVMEGFSN 835

Db 856 QGYANPDVSGYLAVWVPVGAS--DNQDARTAFSTKESGNSAYRTNAAAPDSNVIFEAFSN 913

Qy 836 FOAFATKKEEYTNVIAKNVDFABMGVTDPEMAPQYVSSDTGSLDSVIONGYAFTDRY 895

Db 914 FVYPTKESERANVRIAQNADEFASLGTFSFMAPOYNSSKDRTEFSDTIDNGYAFTDRY 973

Qy 896 DLGISKPNKYGTADLVKAIALKHSKGIKWADWVPDQMYAPEKEVVTATRVDKYGPV 955

Db 974 DLGMEPNKYGTDEDLRNAIQALHAKGLQVWADWPDQIYNLPGEVATVTRVDDRGVW 1033

Qy 956 AGSQITNLYVVDGKSSKQDQAKYGGAFLEBLQAKYFELPARKQISTGVPMVDPVKIKQ 1015

Db 1034 KDAIINNLLYVN--TIGGGEYOKKYGGAFLDKLQKLYPEIFTKQVSTGVAIDPSOKITE 1092

Qy 1016 WSAKYFNGTNIILGRGAGVYLDQATNTYFNISDNKEINFLPKTLNLDQSO--VGRSYDGK 1073

Db 1093 WSAKYFNGTNIILHRSQGVFLKADG-QOYNLGTGTTK-QFLPLQLTGEKKQNGEVKQND 1150

Qy 1074 G-YVYIYSTSGYAKNTFISEG-DKWIYFDNNGYMYVTGAQSN-----GVNYIYFSLNGLQ 1125

Db 1151 GNYIYFDLAGNMVKNFTFIEDSVGNWYFFDQDGKVENKHFDVDSYGEKGYFFLKNQVS 1210

Qy 1126 LRDAILKNEDGTIATYGNDRRYENGYQYFMSGVWRHFNNGEMSVGLTVIDGQVOYFDEM 1185

Db 1211 FRGGLVOTDNGTY-YFDYNGKVRN-----QTINAGAM-----IYTLDEN 1249

Qy 1186 G-----YQAKGKFVTTAD-GKI 1201

Db 1250 GKLIKASYNDSAEYPTSTDVGEM 1272

RESULT 12

US-09-995-749A-2

; Sequence 2, Application US/09995749A

; Patent No. 6867026

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAOUI, HAKIM

; APPLICANT: LEER, ROBERT-JAN

; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES

; FILE REFERENCE: BOA3388-CIP

; CURRENT APPLICATION NUMBER: US/09/995,749A

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: 09/604,957

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: EPO 00201871.1

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 1781

; TYPE: PRT

; ORGANISM: Lactobacillus reuteri

US-09-995-749A-2

Query Match 29.4%; Score 2276; DB 2; Length 1781;

Best Local Similarity 42.4%; Pred. No. 6.9e-144;

Matches 518; Conservative 183; Mismatches 382; Indels 140; Gaps 37;

Qy 35 LVKADSNESKQISNDSNTSVVTANESNVITEATSKEAASSQTN--HTVTTSSSTSV 92

Db 637 IVLVNGKEVXKRLVND-----TKGAAGFRNRDVYKVNPAIENSSM 677

Qy 93 VNPKEVSNPVTV-CETASNGEKLQNTTT-----VDKTSEAAANNISKQTTADTDVD 146

Db 678 SGFGIITLPVTVKENVOLVHRFSNDVKTGEGNVDFWSEL-----MPVKD 724

Qy 147 DSNAAQLILEKLPNVKEIDGKYIYD--NNGKVRNFTLIADGKILHFDGTGAYTDTSD 205

Db 725 SFQKNGPL--KQFGLQTINGQOYIIDPTTGPQRKNFLQSGNNWIYFD-----SDTGVG 777

Qy 206 TVNKDIV-----TTRSN-LYKKNQVYDRSAQSFHVHDHYLTAEBSWYRPKYILKDGKTWQ 260

Db 778 TNALEQFAKGTVSSNEQYRNGNAAYSDDKSIENVNGYLTADTWYRPQKILKDGTTWD 837

Qy 261 STEKDFRPLMTWPDQETORQYVYNNAAQLGINKT-----YDDTSNQLNATAATQAK 316

Db 838 SKETDMRPLMVWPNLTLTQAYLYNMKQHGMLLPALPFFNADADPAELNHYSEIVQON 897

Qy 317 IEAKITTLKNTDMLRTQISAFVKTQSAMNSDSEK-PFDD-HLQNGAVLYDNEGKLTTPYAN 374

Db 898 IEKRISSETGNTDMLRTLMHDFVTNNPMKNKDSENVNFSGIQQGGFLKYEN-SDLTTPYAN 956

Qy 375 SNYRILNRTPNTQTKKDPRYTADNTIGGYEFLLANDVNSNPVQAEOLNHLFLMNFQ 434

Db 957 SDYRLLRGMPIN-----IKQDTYRQGEFLLANDIDNSNPVQAEOLNHLFLMNFQ 1007

Qy 435 NIYANDPDANFDSIRVDADVNDADLLQIAGDYLAAGKIHKNDAANDHLSILEAWSN 494

Db 1008 TITANNDQANFDSVRVDAPDNIDADLNNIAQDYFNAAAYGM-DSDAVSNKHINILEDWNHA 1066

Qy 495 DTPYLLHDDGNNMINMDNKLRLSLFLSLAKPLNORSQGNPLITNSLVNRTDDNAETAAPVS 554

Db 1067 DPEYFNKIGNPOLTWDDTIKNSLNHGLSDATN-RWGLDAIHQSADRENNSTENVVIFN 1125

Qy 555 YSFIKRAHSEVDQLIADIKAIEINPNVVG---YSFTMBEIKKAFIYNKDLATEKKYTH 611

Db 1 LLANDTNSNPVQAEQLNWLHLYMNYGSIANDPEANFDGVRVDAVDNVDNADLLQIAD 60
Qy 467 YLKAAGIHKNDKAANDHLSILEAWSNDPTPYLLHDDGNNINMDNKLRLSLFLSLAKPLN 526
Db 61 YLKAHGVGDSEKNAHLSILEAWSNDPOYNKDTKGALPDKRLSLFLALTRPLE 120
Qy 527 Q-----RSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLIADIKAELN 578
Db 121 KDAASNKNEIRSGLEPVITNSLNNRSAGKNSERMANYPFIRAHDSVQTVIAKIIKAQIN 180
Qy 579 PNWGVGSFTMEIKKAFIYNKDLALATEKYYTHYNTALSYALLTNKSSVPRVYGDMTF 638
Db 181 PKTDGLTFLDELKQAFKTYNEDMRQAKKYYTQSNPTAYALMLSNKDSITRLYYGDMYS 240
Qy 639 DGOQYMAHKTINYEATETLLKARIKYVSGQAMRNOQVGNSE-----IITSVRYCK 689
Db 241 DGOQYMAHKTINYEATETLLKARIKYVSGQAMRNOQVGNSE-----IITSVRYCK 689
Qy 690 GALKAFDTGRTTRTSGVAVIEGNNPSLRKASDRVVMGAAHKNQAYRPLLLTTDNGI 749
Db 301 GANEATDQSEATKTQGMVITSNPSLRKASDRVVMGAAHKNQAYRPLLLTTDNGI 749
Qy 750 KAYHSDQEAAGLVRTNDRGELIFTAADIKG-YANPQVSGYLGWVPVGAALIKMFALRL 808
Db 361 TSYTSDAAKSLYRKNDKGELVFDASDIQGLYNLPQVSG-LAVWVPVGAS--DNQDVRV 417
Qy 809 ARPHQOMAS---VHONAAALDSRMPEGFSNFOAFATKKEEYTNVIAKNVDFPAEWGVT 865
Db 418 AASNKANATQVYESSALDSQIYEGFSNFOAFATKKEEYTNVIAKNVDFPAEWGVT 477
Qy 866 FEMAPQYVSSDGSFLDSVIQNGYAFTRDYLIGSKPNKYGTADDLVKAIKALHSKGIKV 925
Db 478 FEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVKALHSKGIOV 537
Qy 926 MADWVPDQ 933
Db 538 IADWVPDQ 545

RESULT 15
US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-5

Query Match 20.0%; Score 1552; DB 2; Length 523;
Best Local Similarity 58.6%; Pred. No. 5.1e-96;
Matches 311; Conservative 76; Mismatches 132; Indels 12; Gaps 6;
Qy 407 LLANDVNSNPVQAEQLNWLHLYMNYGSIANDPEANFDGVRVDAVDNVDNADLLQIAD 466
Db 1 LLANDVNSNPVQAEQLNWLHLYMNYGSIANDPEANFDGVRVDAVDNVDNADLLQIAD 60
Qy 467 YLKAAGIHKNDKAANDHLSILEAWSNDPTPYLLHDDGNNINMDNKLRLSLFLSLAKPLN 526
Db 61 YLKAHGVGDSEKNAHLSILEAWSNDPOYNKDTKGALPDKRLSLFLALTRPLE 120

Qy 527 QBSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLIADIKAELNPNV- 585
Db 121 INGTWQRFVYVWDRSDSTNEALPNYSFVRAHDSVQTVIAQIV-SOLYDPDENS 179
Qy 586 FTMEIKKAFIYNKDLALATEKYYTHYNTALSYALLTNKSSVPRVYGDMTFDDGQYMA 645
Db 180 PITEQLAAAFKVYNEDEKLADKKYTOYNMASAYAMLLTNKDTVPVYVYGDLYTDDGQYMA 239
Qy 646 HKTINYEATETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKGAALKATDGTDRTRTS 705
Db 240 TKSPYDAINTLLKARVQVAGGQSN---SVDSNDVLTSVRYGKGAALKATDGTDRTRTE 296
Qy 706 GVAVIEGNNPSLRKASDRVVMGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYT 765
Db 297 GGVIVSNNAEQLDEGHTVTLHMGAAHKNQAYRALLSTADGLAYYDTDENAP--VAYT 354
Qy 766 NDRGELIFTAADIKGANPQVSGYLGWVPVGAALIKMFALRLARPHQOMAS---VHON 822
Db 355 DANGDLIFTNESIYGVQNPQVSGYLAWVPVGAQDDQ--DARTASDTTNTSDKVPHSNA 412
Qy 823 ALDSRMPEGFSNFOAFATKKEEYTNVIAKNVDFPAEWGVTDFEMAPQYVSSDGSFLD 882
Db 413 ALDSQVIYEGFSNFOAFATDSSEYTNVIAQNAQDQFQKMGVTSFQLAPQYRSSTDTSF 472
Qy 883 SVIONGYAFTRDYLIGSKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 933
Db 473 SIIONGYAFTRDYLIGSKPNKYGTADDLVKAIKALHSKGIOAIADWVPDQ 523

Search completed: February 11, 2006, 19:45:34
Job time: 38.2941 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 132.022 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-34
Perfect score: 7741
Sequence: 1 MDKKVRYKLRVKRWTVTS.....DRYGRISYYDANSRGRVIRN 1475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7741	100.0	1475	3	US-09-740-274-2
2	7741	100.0	1475	4	US-10-383-930-34
3	7741	100.0	1475	5	US-10-797-821-34
4	5285	68.3	1375	3	US-09-740-274-4
5	5285	68.3	1375	4	US-10-383-930-35
6	5285	68.3	1375	5	US-10-797-821-35
7	4580	59.2	1590	4	US-10-383-930-37
8	4580	59.2	1590	5	US-10-797-821-37
9	3817.5	49.3	1430	3	US-09-740-274-6
10	3817.5	49.3	1430	4	US-10-383-930-36
11	3817.5	49.3	1430	5	US-10-797-821-36
12	3434.5	44.4	1554	4	US-10-383-930-38
13	3434.5	44.4	1554	5	US-10-797-821-38
14	3196.5	41.3	1518	4	US-10-383-930-40
15	3196.5	41.3	1518	5	US-10-797-821-40
16	3161	40.8	1497	5	US-10-484-218-18
17	2945.5	38.1	1365	4	US-10-383-930-39
18	2945.5	38.1	1365	5	US-10-797-821-39
19	2877.5	37.2	1595	5	US-10-484-218-20
20	2537	32.8	2057	4	US-10-417-280A-2
21	2460	31.8	1006	5	US-10-484-218-22
22	2340	30.2	1777	5	US-10-484-218-12
23	2321	30.0	1771	5	US-10-484-218-14
24	2276	29.4	1781	3	US-09-995-749A-2
25	1657	21.4	545	3	US-09-995-749A-10
26	1547.5	20.0	522	3	US-09-995-749A-11
27	1379	17.8	535	3	US-09-995-749A-13

28	1377.5	17.8	787	5	US-10-484-218-16	Sequence 16, Appl
29	1289.5	16.7	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1249.5	16.1	525	5	US-10-484-218-23	Sequence 23, Appl
31	683.5	8.8	224	5	US-10-484-218-4	Sequence 4, Appl
32	673	8.7	223	5	US-10-484-218-6	Sequence 6, Appl
33	647	8.4	223	5	US-10-484-218-10	Sequence 10, Appl
34	577	7.5	221	5	US-10-484-218-2	Sequence 2, Appl
35	552	7.1	221	5	US-10-484-218-8	Sequence 8, Appl
36	456	5.9	2710	4	US-10-011-366-6	Sequence 6, Appl
37	456	5.9	2710	4	US-10-354-774-6	Sequence 6, Appl
38	456	5.9	2710	4	US-10-271-012-6	Sequence 6, Appl
39	456	5.9	2710	4	US-10-729-122-6	Sequence 6, Appl
40	456	5.9	2710	4	US-10-729-039-6	Sequence 6, Appl
41	456	5.9	2710	5	US-10-729-527-6	Sequence 6, Appl
42	456	5.9	2710	5	US-10-727-898-6	Sequence 6, Appl
43	456	5.9	2710	6	US-10-728-696-6	Sequence 6, Appl
44	456	5.9	2710	6	US-11-001-241-6	Sequence 6, Appl
45	402	5.2	811	4	US-10-011-366-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans

Query Match	100.0%	Score	7741;	DB	3;	Length	1475;
Best Local Similarity	100.0%	Pred. No.	0;				
Matches	1475;	Conservative	0;	Mismatches	0;	Indels	0;
		Gaps	0;				
Qy	1	MDKKVRYKLRVKRWTVTSVASAVMTLTTLSSGLVKADSNESKSQISNDSNTSVVTANE	60				
Db	1	MDKKVRYKLRVKRWTVTSVASAVMTLTTLSSGLVKADSNESKSQISNDSNTSVVTANE	60				
Qy	61	ESNVITEATSKQEAASQTNTHTVTTSSSTSVNPKKVVSNPYTVGTETASNGEKLQNTT	120				
Db	61	ESNVITEATSKQEAASQTNTHTVTTSSSTSVNPKKVVSNPYTVGTETASNGEKLQNTT	120				
Qy	121	TVDKTSEAAANNISKQTTTEADTDVDDNSNAANLQLEKLPNVKEIDGKYVYDNNKGVRT	180				
Db	121	TVDKTSEAAANNISKQTTTEADTDVDDNSNAANLQLEKLPNVKEIDGKYVYDNNKGVRT	180				
Qy	181	NFTLIADGKILHFDETGYTDTSDTWNKOIVTTRSNLYKKYQNVYDRSAQSFEHVHYL	240				
Db	181	NFTLIADGKILHFDETGYTDTSDTWNKOIVTTRSNLYKKYQNVYDRSAQSFEHVHYL	240				

Db 181 NPTLIADGKILHFDGTAYDTSIDTVANKD I V T T R S N L Y K K Y N Q V Y D R S A Q S F E H V D H Y L 240

Qy 241 T A E S W Y R P K Y I L K D G K T W T Q S T E K D F R P L L M T W P D Q E T O R Q Y Y N M A Q L G I N K Y T D D T 300

Db 241 T A E S W Y R P K Y I L K D G K T W T Q S T E K D F R P L L M T W P D Q E T O R Q Y Y N M A Q L G I N K Y T D D T 300

Qy 301 S N Q L Q L N T A A A T I O A K I E A K I T T L K N T D W L R O T I S A F V K T O S A W N S D S E K P F D D H L O N G A 360

Db 301 S N Q L Q L N T A A A T I O A K I E A K I T T L K N T D W L R O T I S A F V K T O S A W N S D S E K P F D D H L O N G A 360

Qy 361 V L Y D N E G K L T P Y A N S N Y R I L N R T P N T Q G K D P R Y T A D N T I G G Y E F L L A N D V D N S N P V V Q 420

Db 361 V L Y D N E G K L T P Y A N S N Y R I L N R T P N T Q G K D P R Y T A D N T I G G Y E F L L A N D V D N S N P V V Q 420

Qy 421 A E Q L N W L H F L M F N F G N I Y A N D P A N F D S I R V D A V D N V D A D L L Q I A G D Y L K A A G I H K N D K A 480

Db 421 A E Q L N W L H F L M F N F G N I Y A N D P A N F D S I R V D A V D N V D A D L L Q I A G D Y L K A A G I H K N D K A 480

Qy 481 A N D H L S I L E A W S D N D T P V L H D D G N M I N M D N K L R L S L F S L A K P L N Q R S G M N P L I T N S I V 540

Db 481 A N D H L S I L E A W S D N D T P V L H D D G N M I N M D N K L R L S L F S L A K P L N Q R S G M N P L I T N S I V 540

Qy 541 N R T D N A E T A A V P S Y S F I R A H D S E V Q D L I A D I I K A E I N P N V V G Y S F T M E E I K K A F E I Y N K 600

Db 541 N R T D N A E T A A V P S Y S F I R A H D S E V Q D L I A D I I K A E I N P N V V G Y S F T M E E I K K A F E I Y N K 600

Qy 601 D L L A T E K K Y T H Y N T A L S A L L T N K S S P R V Y Y G M F T D D G Q Y M A H K T I N Y E A I E T L L K A 660

Db 601 D L L A T E K K Y T H Y N T A L S A L L T N K S S P R V Y Y G M F T D D G Q Y M A H K T I N Y E A I E T L L K A 660

Qy 661 R I K Y Y S G G A M N Q V G N S E I T S V R Y G K G A L K A T D T G D R T R T S G V A V I E G N N P S L R L K 720

Db 661 R I K Y Y S G G A M N Q V G N S E I T S V R Y G K G A L K A T D T G D R T R T S G V A V I E G N N P S L R L K 720

Qy 721 A S D R V V N V N G A A H K N Q A Y R P L L L T T D N G I K A Y H S Q E A A G L V R Y T N D R G E L I F T A A D I K G 780

Db 721 A S D R V V N V N G A A H K N Q A Y R P L L L T T D N G I K A Y H S Q E A A G L V R Y T N D R G E L I F T A A D I K G 780

Qy 781 Y A N P O V S G Y L G V W P V G A A L I K M F A L R L A R P H Q Q M A S V H Q N A A L D S R V M F E G S F N Q A P A 840

Db 781 Y A N P O V S G Y L G V W P V G A A L I K M F A L R L A R P H Q Q M A S V H Q N A A L D S R V M F E G S F N Q A P A 840

Qy 841 T K K E E Y T N V I A K N V D K F A E G V T D F E M A P Q V Y S T D G S F L D S V I Q N G Y A F T D R Y D L G I S 900

Db 841 T K K E E Y T N V I A K N V D K F A E G V T D F E M A P Q V Y S T D G S F L D S V I Q N G Y A F T D R Y D L G I S 900

Qy 901 K P N K Y G T A D D L V K A I K A L H S K G I K V M A D W P D M Y A F P E K E V V T A T R V D K Y G T P V A G S Q I 960

Db 901 K P N K Y G T A D D L V K A I K A L H S K G I K V M A D W P D M Y A F P E K E V V T A T R V D K Y G T P V A G S Q I 960

Qy 961 K N T L Y V D G K S G K D Q Q A K Y G A F L E E L O A K Y P E L F A R K Q I S T G V P M D P S V K I K O W S A K Y 1020

Db 961 K N T L Y V D G K S G K D Q Q A K Y G A F L E E L O A K Y P E L F A R K Q I S T G V P M D P S V K I K O W S A K Y 1020

Qy 1021 F N G T N I L G R G A G Y L K D Q A T N Y F N I S N K E I N F L P K T L L N O D S O V G F S Y D G K G V Y Y Y S T 1080

Db 1021 F N G T N I L G R G A G Y L K D Q A T N Y F N I S N K E I N F L P K T L L N O D S O V G F S Y D G K G V Y Y Y S T 1080

Qy 1081 S G Y Q A K N T F I S E G D K W Y Y F D N N G Y N V T G A Q S I N G V N Y Y F L S N G L Q R D A I L K N E D G T Y A Y 1140

Db 1081 S G Y Q A K N T F I S E G D K W Y Y F D N N G Y N V T G A Q S I N G V N Y Y F L S N G L Q R D A I L K N E D G T Y A Y 1140

Qy 1141 Y G N D G R R Y E N G Y Y Q P M S G W R H F N N G E M S V G L T V I D G V Q V Y P D E M G Y Q A K G F V T T A D G K 1200

Db 1141 Y G N D G R R Y E N G Y Y Q P M S G W R H F N N G E M S V G L T V I D G V Q V Y P D E M G Y Q A K G F V T T A D G K 1200

Qy 1201 I R Y F D K O S G N M Y R N F I E N E E G K W L Y L G E D G A A V T G S O T I N G Q H L Y F R A N G V Q V K G E F V T 1260

Db 1201 I R Y F D K O S G N M Y R N F I E N E E G K W L Y L G E D G A A V T G S O T I N G Q H L Y F R A N G V Q V K G E F V T 1260

Qy 1261 D H H G R I S Y Y D N S G D Q I R N R F V R N A Q O G W F Y D N N G Y A V T G A R T I N G Q L L Y F R A N G V Q V K 1320

Db 1261 D H H G R I S Y Y D N S G D Q I R N R F V R N A Q O G W F Y D N N G Y A V T G A R T I N G Q L L Y F R A N G V Q V K 1320

RESULT 2

US-10-383-930-34
; Sequence 34, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-34

Query Match 100.0%; Score 7741; DB 4; Length 1475;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 M D K K V R Y K L R V K K R W T V S V A S A V M T L T L S G G L V K A D S N E S K S Q I S N D S N T S V V T A N E 60

Db 1 M D K K V R Y K L R V K K R W T V S V A S A V M T L T L S G G L V K A D S N E S K S Q I S N D S N T S V V T A N E 60

Qy 61 E S N V I T E A T S K Q E A A S S Q T N H T V T T S S S T S V N P K E Y V S N P Y T V G E T A S N G E K L Q N Q T T 120

Db 61 E S N V I T E A T S K Q E A A S S Q T N H T V T T S S S T S V N P K E Y V S N P Y T V G E T A S N G E K L Q N Q T T 120

Qy 121 T V D K T S E A A A N N I S K Q T T E A D T D V I D D S N A A N L Q I L E K L P N V K E I D G K Y Y Y D N N G K V R T 180

Db 121 T V D K T S E A A A N N I S K Q T T E A D T D V I D D S N A A N L Q I L E K L P N V K E I D G K Y Y Y D N N G K V R T 180

Qy 181 N F T L I A D G K I L H F D E T G A Y T D T S I D T V N K D I V T T R S N L Y K K Y N Q V Y D R S A Q S F E H V D H Y L 240

Db 181 N F T L I A D G K I L H F D E T G A Y T D T S I D T V N K D I V T T R S N L Y K K Y N Q V Y D R S A Q S F E H V D H Y L 240

Qy 241 T A E S W Y R P K Y I L K D G K T W T Q S T E K D F R P L L M T W P D Q E T O R Q Y Y N M A Q L G I N K Y T D D T 300

Db 241 T A E S W Y R P K Y I L K D G K T W T Q S T E K D F R P L L M T W P D Q E T O R Q Y Y N M A Q L G I N K Y T D D T 300

Qy 301 S N Q L Q L N T A A A T I O A K I E A K I T T L K N T D W L R O T I S A F V K T O S A W N S D S E K P F D D H L O N G A 360

Db 301 S N Q L Q L N T A A A T I O A K I E A K I T T L K N T D W L R O T I S A F V K T O S A W N S D S E K P F D D H L O N G A 360

Qy 361 V L Y D N E G K L T P Y A N S N Y R I L N R T P N T Q G K D P R Y T A D N T I G G Y E F L L A N D V D N S N P V V Q 420

Db 361 V L Y D N E G K L T P Y A N S N Y R I L N R T P N T Q G K D P R Y T A D N T I G G Y E F L L A N D V D N S N P V V Q 420

Qy 421 A E Q L N W L H F L M F N F G N I Y A N D P A N F D S I R V D A V D N V D A D L L Q I A G D Y L K A A G I H K N D K A 480

Db 421 A E Q L N W L H F L M F N F G N I Y A N D P A N F D S I R V D A V D N V D A D L L Q I A G D Y L K A A G I H K N D K A 480

```
QY 481 ANDHLSILEAWSNDTPYLHDDGDNMNMNDKRLSLFSLAKPLNQRSGMPLTNSLV 540
Db 481 ANDHLSILEAWSNDTPYLHDDGDNMNMNDKRLSLFSLAKPLNQRSGMPLTNSLV 540
QY 541 NRTDDNAETAAPVPSYFIRAHSEVDLIADIKAIEINPNVGYSTMEIEIKAEIYNK 600
Db 541 NRTDDNAETAAPVPSYFIRAHSEVDLIADIKAIEINPNVGYSTMEIEIKAEIYNK 600
QY 601 DLLATEKKYTHYNTALSVALLTNKSVPVRYGDMFTDDGOVMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKYTHYNTALSVALLTNKSVPVRYGDMFTDDGOVMAHKTINYEAIETLLKA 660
QY 661 RIKYVSGQAMRNQVGNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNPRLRK 720
Db 661 RIKYVSGQAMRNQVGNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNPRLRK 720
QY 721 ASDRVVNVNNGAAHQAQVPELITTDNGIKAVHSQDEAAGLVRYTNDRGELIFTAADIKG 780
Db 721 ASDRVVNVNNGAAHQAQVPELITTDNGIKAVHSQDEAAGLVRYTNDRGELIFTAADIKG 780
QY 781 YANPOVSGYLVGVVPGAALIKFALRLARPHQQMASVHONAAALDSRVNMFEGSFQAPA 840
Db 781 YANPOVSGYLVGVVPGAALIKFALRLARPHQQMASVHONAAALDSRVNMFEGSFQAPA 840
QY 841 TKKEEYTNVVIKQVNDKFAEWGVTDFEMAPQVVSSTGDSFLDSVIQNGYAFTRDYDLGIS 900
Db 841 TKKEEYTNVVIKQVNDKFAEWGVTDFEMAPQVVSSTGDSFLDSVIQNGYAFTRDYDLGIS 900
QY 901 KENKYGTADDLYKAIKALHSKIKIWMADVPQMAFPEKEVVTATRVDKYGTTPVAGSQI 960
Db 901 KENKYGTADDLYKAIKALHSKIKIWMADVPQMAFPEKEVVTATRVDKYGTTPVAGSQI 960
QY 961 KNTLYVVDGSKGQDOQAQYGAFLLELOAKYPELPARKQISTGVPMDSVVKI KOWSAKY 1020
Db 961 KNTLYVVDGSKGQDOQAQYGAFLLELOAKYPELPARKQISTGVPMDSVVKI KOWSAKY 1020
QY 1021 FNGTNILGRGAGYVLKQDQATNTYFNISDNKEINFLPKTLNLDQSDQVGSYDGKGVYVYST 1080
Db 1021 FNGTNILGRGAGYVLKQDQATNTYFNISDNKEINFLPKTLNLDQSDQVGSYDGKGVYVYST 1080
QY 1081 SGYQAKNTFISEGDKWYYPDNNGYVMTGAQSIQNGVNYFPLSNGLOLRDAILKNEDGTAY 1140
Db 1081 SGYQAKNTFISEGDKWYYPDNNGYVMTGAQSIQNGVNYFPLSNGLOLRDAILKNEDGTAY 1140
QY 1141 YGNDGRRYENGYYQFMGSQVWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTDGK 1200
Db 1141 YGNDGRRYENGYYQFMGSQVWRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTVTDGK 1200
QY 1201 IRYFDKQSGNMYRNRFIENEKGWLYLGEDGAAVTGSQTINGQHLYFRANGVQVKGFEVT 1260
Db 1201 IRYFDKQSGNMYRNRFIENEKGWLYLGEDGAAVTGSQTINGQHLYFRANGVQVKGFEVT 1260
QY 1261 DHGGRISYDGNSGDQIRNRFVRNAQGWYFDPNNGYAVTGARTINGQLLYFRANGVQVK 1320
Db 1261 DHGGRISYDGNSGDQIRNRFVRNAQGWYFDPNNGYAVTGARTINGQLLYFRANGVQVK 1320
QY 1321 GBFVTDYGRISYDGNSGDQIRNRFVRNAQGWYFDPNNGYAVTGARTINGQHLYFRAN 1380
Db 1321 GBFVTDYGRISYDGNSGDQIRNRFVRNAQGWYFDPNNGYAVTGARTINGQHLYFRAN 1380
QY 1381 GVQVKGFEVTDHGRISYDGNSGDQIRNRFVRNAQGWYFDPNNGYAVTGARTINGQHL 1440
Db 1381 GVQVKGFEVTDHGRISYDGNSGDQIRNRFVRNAQGWYFDPNNGYAVTGARTINGQHL 1440
QY 1441 YFRANGVQVKGFEVTDYGRISYDANSGERVRIN 1475
Db 1441 YFRANGVQVKGFEVTDYGRISYDANSGERVRIN 1475
```

RESULT 3

US-10-797-821-34

; Sequence 34, Application US/10797821

; Publication No. US20050031633A1

```
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-10-797-821-34
```

Query Match 100.0%; Score 7741; DB 5; Length 1475;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MDKVVYKLRKVKRWTVSVASAVMTLTTLTSGGLVKADSNESKQISNDSNTSVVTANE 60
Db 1 MDKVVYKLRKVKRWTVSVASAVMTLTTLTSGGLVKADSNESKQISNDSNTSVVTANE 60
QY 61 ESNVITEATSKQEAASSQTNHTVTTSSTSVNPKVSVNPTVGTASNGEKLQNTT 120
Db 61 ESNVITEATSKQEAASSQTNHTVTTSSTSVNPKVSVNPTVGTASNGEKLQNTT 120
QY 121 TVDKTSEAAANNISKQTTTEADTVLDDSNAAANLQLEKLPNVKEIDGKYIYDNNKQVRT 180
Db 121 TVDKTSEAAANNISKQTTTEADTVLDDSNAAANLQLEKLPNVKEIDGKYIYDNNKQVRT 180
QY 181 NPTLIADGKILHFDETGAYTDTSDTVNKKDITVTRSNLYKKYQVYVDRSAQSFHVHDYVL 240
Db 181 NPTLIADGKILHFDETGAYTDTSDTVNKKDITVTRSNLYKKYQVYVDRSAQSFHVHDYVL 240
QY 241 TAESWYRPKYILKDGKWTQSTEKDFRPLMTWMPDQETQRYVNMNAQLGINKTYDDT 300
Db 241 TAESWYRPKYILKDGKWTQSTEKDFRPLMTWMPDQETQRYVNMNAQLGINKTYDDT 300
QY 301 SNQQLNLTAAATIQAIEAKITTLKNTDWLRQTTSFVKTQSAMNSDSEKPPDDHLONGA 360
Db 301 SNQQLNLTAAATIQAIEAKITTLKNTDWLRQTTSFVKTQSAMNSDSEKPPDDHLONGA 360
QY 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTGKDPRTVADNTIGGYEFLANDVDNPNPVQ 420
Db 361 VLYDNEGKLTYPYANSNYRILNRTPTNQTGKDPRTVADNTIGGYEFLANDVDNPNPVQ 420
QY 421 AEQLNWLHFLMNFNIYANDPDANFDSIRVDAVNVDADLLQIAGDYLKAAKGIHKNDKA 480
Db 421 AEQLNWLHFLMNFNIYANDPDANFDSIRVDAVNVDADLLQIAGDYLKAAKGIHKNDKA 480
QY 481 ANDHLSILEAWSNDTPYLHDDGDNMNMNDKRLSLFSLAKPLNQRSGMPLTNSLV 540
Db 481 ANDHLSILEAWSNDTPYLHDDGDNMNMNDKRLSLFSLAKPLNQRSGMPLTNSLV 540
QY 541 NRTDDNAETAAPVPSYFIRAHSEVDLIADIKAIEINPNVGYSTMEIEIKAEIYNK 600
Db 541 NRTDDNAETAAPVPSYFIRAHSEVDLIADIKAIEINPNVGYSTMEIEIKAEIYNK 600
QY 601 DLLATEKKYTHYNTALSVALLTNKSVPVRYGDMFTDDGOVMAHKTINYEAIETLLKA 660
Db 601 DLLATEKKYTHYNTALSVALLTNKSVPVRYGDMFTDDGOVMAHKTINYEAIETLLKA 660
```


Db 1130 DNNGYMTGAQINGANYFLSLNGQLRANAIYDNGKNVLSYVGNDRRYENGYYLF-CQQ 1188
QY 1160 WRHFNNGEMSVGLTWIDGOVYFDEMGOAKGKFTTTADGKIRYFDKSGNNMYRNFTE 1219
Db 1189 WRYFQNGIMAVGLTRVHGAVQYFDASGFOAKGQFITTADGKLYRFDSDGNQISNRFVRN 1248
QY 1220 BEGKWLILGEDCAATGSGTNGOHLYPFRANGVOVKGEPVTDHGHGISYDNGSDQIRN 1279
Db 1249 SKGEWFLFDHNGVAVTGTVTFGQRLYFKPNGVOAKGEFIRDANGLYRYDPSNGNEVRN 1308
QY 1280 RFRVNAQGWYFDDNNGYAVTGARTING-----QLLYER 1313
Db 1309 RFRVNSKGWFLFDHNGIAVTGARVNVNGHASILSLMVER 1347

RESULT 6

US-10-797-821-35
; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PR1
; ORGANISM: Streptococcus mutans
US-10-797-821-35

Query Match 68.3%; Score 5285; DB 5; Length 1375;
Best Local Similarity 75.9%; Pred. No. 1.2e-296;
Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;
QY 1 MDKVRYLKRVKRWTVSVASAVMTLTLSGGLVKADSNESKQISNDSNTSVVTANE 60
Db 1 MEKVRFLKRVKRWTVSVIASAVVTLTSLGSLVKADSTDDRQAVTESQASLVTTSE 60
QY 61 -----ESNVITEATSKQEAASQTNHTVTSSSTSVVNPKEVSNPYTVCETA-- 109
Db 61 AAKETLTATDTSATSSQPTA--TVTDNVSTTQSTNTTANTANFVVKPTTTSEQAKT 118
QY 110 SNGEKLOQTTVTDKTSSEA-----AANNISKOTTEAD-----TDVIDSN 149
Db 119 DNSDKIITTSKAVNRLTATGKVPANNNTAHPKTVTDKIVPKPKIGKLKQPSLSQDDI 178
QY 150 AANLQILEKLPNKEIDGKYIYDNGKVRNTFTLIADGKILHFDDETGAYTDTSIDTVNK 209
Db 179 AA----LGNVKIRKGVKYYKEDGTLQXNALNINGKTFPFDGALSNNTLPSKKG 234
QY 210 DIVTT-RNLKYKNQVDRSQAQSEHVDHLYLTAEWSRPKYLKDGKTTQSTKDFRP 268
Db 235 NITNDNTNSFAQYQNVYSTDVANFEHVDHLYLTAEWSRPKYLKDGKTTQSTKDFRP 294
QY 269 LLMTWMPDQETORQVYVNYMNAQLGHTQYNTATSTPLQLNLAQTIQTKEEKIITAKNTN 328

Db 295 LLMTWMPDQETORQVYVNYMNAQLGHTQYNTATSTPLQLNLAQTIQTKEEKIITAKNTN 354
QY 329 WLRTISAFVTKQSAWNSSEKPFDDHLQNGAVLYDNEGKLTTPYANSYRILNRTPTNOT 388
Db 355 WLRTISAFVTKQSAWNSSEKPFDDHLQNGAVLYDNEGKLTTPYANSYRILNRTPTNOT 414
QY 389 GKQDPRTYADNTIGGYEFLLANDVNSNPVQAEOLNWLHFLMNFNGNIYANDPDANFDSI 448
Db 415 GKQDPRTYADNTIGGYEFLLANDVNSNPVQAEOLNWLHFLMNFNGNIYANDPDANFDSI 474
QY 449 RVDADVNDADLLQIAGDYLKAAKGIHKNDKAAANDHLSILEAWSNDTPYLHDDGDNMIN 508
Db 475 RVDADVNDADLLQIAGDYLKAAKGIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNMIN 534
QY 509 MDNKLRLSLFLSLAKPLNQRSGMNPILIINSLVNRITDNDNAETAAPVPSYFIRAHDEVDL 568
Db 535 MDNKLRLSLFLSLAKPLNQRSGMNPILIINSLVNRITDNDNAETAAPVPSYFIRAHDEVDL 594
QY 569 IADIIKAEINPNVVGYSFTMBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKSSV 628
Db 595 IRNIIIRTEINPNVVGYSFTMBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKSSV 654
QY 629 PRVYGDNFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAMRNQQVGNSEIITSVRYG 688
Db 655 PRVYGDNFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAMRNQQVGNSEIITSVRYG 714
QY 689 KGALKATDGTGRTTTSQVAVIEGNNPSLRUKASDRVVVNMGAHQAQYRPLLLTTONG 748
Db 715 KGALKATDGTGRTTTSQVAVIEGNNPSLRUKASDRVVVNMGAHQAQYRPLLLTTONG 774
QY 749 IKAYHSDQEAAGLVRYTNDRGELIFTAADIIGYANPOVSGYLGWVVPVGAALIKMFAURL 808
Db 775 IKAYHSDQEAAGLVRYTNDRGELIFTAADIIGYANPOVSGYLGWVVPVGAALIKMFAURL 832
QY 809 A---RPHQOMASVHQNAALDSRVMPGFSNFQAFATKKEEYTNVIAKNVDKFAEWGVD 865
Db 833 AASTAPSTDGSKSVHQNAALDSRVMPGFSNFQAFATKKEEYTNVIAKNVDKFAEWGVD 892
QY 866 FEMAPQYVSSTDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGIKV 925
Db 893 FEMAPQYVSSTDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGIKV 952
QY 926 MADVVPDQMYAFPEKEVVTATRVDKYGTVPAGSQIKNTLYVVDGKSGKQOQKYGGAFL 985
Db 953 MADVVPDQMYAFPEKEVVTATRVDKYGTVPAGSQIKNTLYVVDGKSGKQOQKYGGAFL 1012
QY 986 EELQAKYPELPARKQISTGVPMPSVKIKQWSAKYFNGTNIILGRGAGVYLKQOQATNTYFN 1045
Db 1013 EELQAKYPELPARKQISTGVPMPSVKIKQWSAKYFNGTNIILGRGAGVYLKQOQATNTYFN 1072
QY 1046 -ISDNKEINFLPKTLN-----QDSQVGSYDGGKGYVYYSYSGYQAKWTFISEGDKWYF 1099
Db 1073 LVSDN---TFLPKSLVNPNNHGTSSSVTGLVPDGGYVYYSYSGYQAKWTFISEGDKWYF 1129
QY 1100 DNNGYMTGAQINGANYFLSLNGQLRANAIYDNGKNVLSYVGNDRRYENGYYLF-CQQ 1159
Db 1130 DNNGYMTGAQINGANYFLSLNGQLRANAIYDNGKNVLSYVGNDRRYENGYYLF-CQQ 1188
QY 1160 WRHFNNGEMSVGLTWIDGOVYFDEMGOAKGKFTTTADGKIRYFDKSGNNMYRNFTE 1219
Db 1189 WRYFQNGIMAVGLTRVHGAVQYFDASGFOAKGQFITTADGKLYRFDSDGNQISNRFVRN 1248
QY 1220 BEGKWLILGEDCAATGSGTNGOHLYPFRANGVOVKGEPVTDHGHGISYDNGSDQIRN 1279
Db 1249 SKGEWFLFDHNGVAVTGTVTFGQRLYFKPNGVOAKGEFIRDANGLYRYDPSNGNEVRN 1308
QY 1280 RFRVNAQGWYFDDNNGYAVTGARTING-----QLLYER 1313
Db 1309 RFRVNSKGWFLFDHNGIAVTGARVNVNGHASILSLMVER 1347

RESULT 7

US-10-383-930-37

QY	706	GVAVTEGNNPSLRLLKASDRVVVVNMGAAHQQAQYRPLLLTTDNGIKAYHSQDEA--AGLVR	763
DB	702	GVGVVMGQNPFSLDGK--VVALNMGAHAHANQYRPLVSTKDGVTATDADASKAGLAVK	760
QY	764	YTNDRGELIFTAADIKGVANPQVSGYLGWVWVPGAA--LTKMFALRLARPHQOQMASVHQ	820
DB	761	RTDENGILYFLNDDLLKGVANPQVSGFLQVWVPGAADDQDIRVAASSTADTGK--SLHQ	818
QY	821	NAALDSRVNPFSGNFQAFATKKEEYTNVVIKAKVDFABMGVTDTFEMAPQYVSSSTDGSGF	880
DB	819	DAAMDSRVNPFSGNFQSFATKKEEYTNVVIKAKVDFABMGVTDTFEMAPQYVSSSTDGSGF	878
QY	881	LDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKVIMADWVPDQWYAPPEK	940
DB	879	LDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQWYTPPKQ	938
QY	941	EVVTRATRDYKGTVPVAGSOIKNTLYVVDGSSGQQAQYKGGAFLEQLQAKYPELPARKQ	1000
DB	939	EVVTRTRTDKFKPAGSQINHSLYVTDTKSSGGDYQAKYGGAFLEQLQAKYPELPARKQ	998
QY	1001	ISTGVPMDFSVKIKQWSAKYFNGTNILGRGAGYVLKXQQAATNTYFNISDNKEINFLPKTLL	1060
DB	999	ISTGQAIDPSVKIKQWSAKYFNGSNILGRGADYVLSDVSNKYFNVA--SDTLFLPSSLL	1056
QY	1061	NQDSQVGSYDGKGVYVYST--SGYQAKNTFISEBQKQYVFDNNGYVMTGAQSGINGVNYF	1119
DB	1057	GRVBSGIRYDGKGIYNSSATGQVKASFITEAGNLVYFKDQYVMTGAQTINGANYFF	1116
QY	1120	LSNGIQLRDALIKNEDGTAYVYGNDRGRYNGYVQFMSGVWRPNNGEMSVGLTVTDGQV	1179
DB	1117	LENGTALRNTLYTDAQNSHYVANDGRYNGYQGF--GNDWRYFKQGNMVAULTVVDGNV	1175
QY	1180	QYFDEMGVQAQKGFVTTADGKIRYFQKQSGMNTNRFEIENEGKWLVLGEGDGAATVGSQT	1239
DB	1176	QYFQDKGVQAQKLIIVTDEGKVRVYFDQHNGNAATNTFIADKTGHVYLGKGVAVTGAQT	1235
QY	1240	INGQHLVFRANGVQVKGFEVTDHHRISYIDGNSGDOIENRFRVNAQGVYFDNNGYAV	1299
DB	1236	VGKQKLYPEANGQVKGDFVTSDEGLKYFYDVSDDMTDTFTIEDKAGNFFYLKDKGAAV	1295
QY	1300	TCARTINGOLLYFRANGVQVKGFEVTDYGRISYIDGNSGDOIENRFRVNAQGVYFDN	1359
DB	1296	TGAQTFIRQKLYFRANGVQVKGFDIVKGTGDKIRYDAKSGEQVFNKTVKAADGKTYVIGN	1355
QY	1360	NGYAV-----TGARTI	1370
DB	1356	DGVAVDPVSVKQTFKQASGALRFYNLKGQLVTGSGHYETANHDWYIQQSKALTBQETI	1415
QY	1371	NGQHLVFRANGVQVKGFEVTDHHRISYIDGNSGDOIENRFRV-----	1412
DB	1416	NGQHLVFKEDGHQVKGQLVTGTDGKVRYYDANSQDQAFNKSVTWNGKTYVYFGNDGTAQT	1475
QY	1413	-----RNAQGVYFDNNGYAVTGAQTINGQHLVYF	1442
DB	1476	GNPKGQTFKQGSIDIRFYSMEGQLVTGSGWYENAGQWLYV--KNGKVLTLGLQTVGSGRVYF	1534
QY	1443	RANGVQVKGFEVTDYGRISYIDANSGERV	1472
DB	1535	DENGLQAKGVARTSDGKIRYFDENSGSMI	1564
RESULT 8			
US-10-797-821-37			
; Sequence 37, Application US/10797821			
; Publication No. US20050031633A1			
; GENERAL INFORMATION:			
; APPLICANT: Smith, Daniel J.			
; APPLICANT: Taubman, Martin A.			
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogen			
; FILE REFERENCE: 25669-020			
; CURRENT APPLICATION NUMBER: US/10/797,821			
; CURRENT FILING DATE: 2004-03-09			
; PRIOR APPLICATION NUMBER: 10/383,930			

```
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRF
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match          59.2%; Score 4580; DB 5; Length 1590;
Best Local Similarity 56.4%; Pred. No. 8.4e-256;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;

QY 1 MDKVRYYKLRKYKRWVTVSVASAVMTLTLTSLGGVLKADSNESKQISNDNNTSVVTANE 60
DB 1 MEKNVRFKMKHKKRWVTVSVASATMLASALGASVASADTDTA-----SDSSNQAVVTGDQ 56

QY 61 ESNVITEATSKOEAAASSQTNHTVTTSSSSTS VVNPK-----VVSNPYTVGETASNGEKL- 115
DB 57 TTN--NQATDQTSIAATSEQASTDAATDQASAAEQOTQGTASTDTAAQTITNANEAK 114

QY 116 -----QNTTVDKTSBAANNISQOTTEADTVDDSNAAANLOILEKLPNVKEIDGKY 169
DB 115 WYPTENENQGFDEMLAE-----KNVATAESDSIPSD-----LAKMSNVKQVDGKY 161

QY 170 YYDNNKGVRTNFTLIADGKILHFDGTAYTDTSDITNKK--DIYVTRSNLYKKXNOYVD 227
DB 162 YYDQDGNVKNFAVSGDKIYYFDETGA YKDTSKVDADKSSASVQNAITFAANRAYS 221

QY 228 RSAQFEHVDHYLTAAESVVRPKYILKDGKWTQSTKDFRPLMTWPDQETQROYVNYM 287
DB 222 TSAKNFEAVNLTADSVVRPKSILKDGKWTESGKDFRPLLMAMWPDTEKRYVNYM 281

QY 288 NALQINKTYDTSNQLQNTAAATIOAKIEKITLKNKDWLROTISAFVKTQSAWNSD 347
DB 282 NKVWIDKTYTAETSQADLTAAAEVLQARIEQKITSENNTKWLREAI SAFVKTQPNWGE 341

QY 348 SEKPPDDHLQNGAVLYDNEGKLTTPYANSNYRILNKTPTNQTGKDPRTY--ADNTIGGYE 405
DB 342 SEKPYDDHLQNGALLFDNQTDLTPTQSNRYRLNKTPTNQTGSLDSRFTYPNPDPLGGYD 401

QY 406 FLLANDVNSNPVQAEQLNMLHFLMNFENIYANDPDANFDSIRVDAVDNVDADLLQIAG 465
DB 402 FLLANDVNSNPVQAEQLNMLHYLLNFGSIYANDADANFDSIRVDAVDNVDADLLQI1SS 461

QY 466 DYLKAAKGIHKNDKAANDHLSILEAWSNDTPYLHDDGDNMINMKNKURLSLFLSLAKPL 525
DB 462 DYLKAAAYGIDKNNKNVANNHVSIVEAWSNDTPYLHDDGDNLMNMNDNKFRLSMLWSLAKPL 521

QY 526 NORSQMNPLITNSLVNRTDQNAETAAPVPSYSFIRAHDSFVODLIADIILKAEINPNVGVYS 585
DB 522 DKRSGLLPLHNSLVNREVDREVETVPSYSPARAHDSFVODIIRDIILKAEINPNFSFGYS 581

QY 586 FTMEIKKAFETYNKDLLATEKKYTHYNTALSYALLTNKGSVPRVYVYCDMFTDDGQYMA 645
DB 582 FTQEEIEQAFKIYNEDLAKTKKYTHYNVPLSYTLTLLTNKGSIPRVYVYCDMFTDDGQYMA 641

QY 646 HKTINYEAIETLLKARIKYVSGQAMRNQOVGNSIIITSVRYGKALKATDTGDRTRTTS 705
DB 642 NKTNYDAIESLLKARMKYVSGQAMQNYQIGNGEILT SVRYGKALKQSGKGDATTRTS 701

QY 706 GVAVIEGNNPSLRKASDRVVVNMCAAHKNQAYRPLLLTDTONGIKAYHSDQEA--AGLVR 763
```

```
DB 702 GGVVMGNQPNFSLDGK--VVALNMGAAHANQEYRALMYSTKDG VATYATDADASAKGLVK 760
QY 764 YTNDRGELIFTAADIKGYANPOVSGYLGWVPVPGAA---LIMKFAIRLARPHQOQWASVHQ 820
DB 761 RTDENGILYFLNDLDLKGVAANPQVSGFLQWVPVGAADQDIRVAASDASTDGK--SLHQ 818
QY 821 NAAIDSRVMPEFGFSNFQAFATKCKEYTNVITAKNVNKEAEWGVTD FENAPQVVSSTDSGF 880
DB 819 DAAMD SRVMEFGFSNFQAFATKCKEYTNVITAKNVNKEAEWGVTD FENAPQVVSSTDSGF 878
QY 881 LDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAKJALHSHKGIKVMADWVPDQMTAFPEK 940
DB 879 LDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAKJALHSHKGIKVMADWVPDQMTAFPEK 938
QY 941 EVVTATRVDKYGT PVAGSQIKNTLVVVDGKSGKDDQQA KYGGAFLEELQAKYPELFAKQ 1000
DB 939 EVVTATRVDKYGT PVAGSQIKNTLVVVDGKSGKDDQQA KYGGAFLEELQAKYPELFAKQ 998
QY 1001 ISTGVPMDPSPVKIKQWSAKYFNGTNILGRGAGYVLKQDQATNTYFNI SDNKEINFLPKTL 1060
DB 999 ISTGVPMDPSPVKIKQWSAKYFNGTNILGRGAGYVLKQDQATNTYFNI SDNKEINFLPKTL 1056
QY 1061 NODSOVGFSGYDGKGVVYVYST--SGYQAKNTFTISEGDKWYFDMNGYMTVGAQSGINGVNYF 1119
DB 1057 GKVESGIRYDGKGYIYNSSATGQVQKASFTIAGNLTYFGKDGVTWGTGAQTINGANVFF 1116
QY 1120 LNSGLQLRDAI LKNEDEGTYYAYGNDGRYRNGYQYFMSGVWRHFNNGEMSVGLTVIDGOV 1179
DB 1117 LENGTALRNTIYTDAGNSHYVANDGKRYENGYQQF--GNDWRYFPKDGNAVGLTVDGNV 1175
QY 1180 QYFDEMGYQAKGFVTTADGKIRYFDKOSGNMYRNFIEENEGKWL YLGEDGAATVGSQT 1239
DB 1176 QYFDEMGYQAKGFVTTADGKIRYFDKOSGNMYRNFIEENEGKWL YLGEDGAATVGSQT 1235
QY 1240 INGQHLVFRANGVQVKGFEVTDHGRISYSDGNSGDIQIRNRFVRNAOQOWFVFDNNGYAV 1299
DB 1236 VGKQLLFEANGQVKGDFVTSDEGLYFYDVSDDMTDTFIEDKAGNFWFLGDKGA 1295
QY 1300 TGARTINGQLLYFRANGVQVKGFEVTDHGRISYSDGNSGDIQIRNRFVRNAOQOWFVFDN 1359
DB 1296 TGAQTIRGOKLYFRANGVQVKGDIKVGTDGKIRYDAKSGEQVFNKTVKAADGKTVIGN 1355
QY 1360 NGYAV-----TGARTI 1370
DB 1356 DGAVAVPSVVKGTQFKDASGALR FYNLKGQLVTGSGWYETANHDWYI TQSGKALTGEQTI 1415
QY 1371 NGQHLVFRANGVQVKGFEVTDHGRISYSDGNSGDIQIRNRFV----- 1412
DB 1416 NGQHLVFKEDGHQVKGQLVTGTDGKVRYYDANSQDQAFNKS VTYNGKTYYPGNDGTAQTA 1475
QY 1413 -----RNAOQOWFVFDNNGYAVVTGARTINGQHL YF 1442
DB 1476 GNPKGQTPKDGSDIRFYSMEGQLVTGSGWYENAOQOWLYV--KNGKVLTLGLOTVGSORVYF 1534
QY 1443 RANGVQVKGFEVTDHGRISYSDGNSGDIQIRNRFV----- 1472
DB 1535 DENGIOAKGKAVRTSDGKIRYFDENSGSMI 1564

RESULT 9
US-03-740-274-6
; Sequence 6, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
```

```
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRP
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match      49.3%; Score 3817.5; DB 3; Length 1430;
Best Local Similarity 52.9%; Pred. No. 9e-212;
Matches 762; Conservative 231; Mismatches 394; Indels 53; Gaps 22;

QY 1 MDKVRYYKLRYKRWVTVSVASAVMTL--TTLGGGLVKADSNESKS-----QISNDSNTS 54
DB 1 METKRRYKRWKRWVTVSVASAVMTL--TTLGGGLVKADSNESKS-----QISNDSNTS 60
QY 55 VVTANEESNVITEATS KQBAASSQTNHHTVTTSSSSTSVVNPKEVSNPYTVGSETASNGEK 114
DB 61 SESSQTDAPKTKQAOTEQTOAQSOAN--VADTSTSTIKETPSQNIITTOANSDDKTVTNYS 119
QY 115 LQNTTIV--DKTSEAAANNISKOT--TEADTDVIDSNAA-----NLQILEKLPVNYKEIDG 167
DB 120 BEAQTSEERTKQSEEAQTASSQALTKQAKELTKQRTAAQENKPNVDLAAIPNVKQIDG 179
QY 168 KYYYDNNCKVPTNLTADGKILHFE--TGAYTDTSDTNDKDIIVTTRSNLYKKYNQYV 226
DB 180 KYYYIGSDGQPKKNFALTNNVKYFVKNTGALTDTTS--QYQFKGLTTLNNDYTPHNOIV 238
QY 227 DRSAQSEFHVHLYTAESWYRPKYILKDGKWTQSTQTEKDFRLLMTWMPDQSTQRYVNY 286
DB 239 NFENTSLETIDNYVADSWYRPKILKNGKWTWASEDLRPLMSWPFDKQTQIAYLNY 298
QY 287 MNAQ--LGINKTVDDTSNQLNIAAATIQAKIEAKITTLKNTDMLRQTIASFVKTOSAWN 345
DB 299 MNQGLGTGENTADSSQESLNLAQTQVKIETKISQTOQTQWLDRDIINSFVKTOPNNW 358
QY 346 SDSEKPFDD----DHLONGAVLYDNSEGLTPYANSYRILNRPPTNQTGKDKDPRYTADNTI 401
DB 359 SQTESDTSAGEKHLOGGGALLYSNDK--TAYANSYRILNRPPTNQTGK--PKYFEDNSS 415
QY 402 GGYEFLLANDVDSNPVQAEQLNWLFLNFGNTIYANDPDANFDSIRVDAVDNVDADLL 461
DB 416 GGYDFLLANDIINSPVQAEQLNWLHLYMNGSIVANDPEANFGVRVDAVDNVDADLL 475
QY 462 QIAGDYLKAAGIHNKKAANDHLSILEAWSNDPTPYLHDDGDNMINNDKRLSLLSFL 521
DB 476 QIASDYLKAHYGVDSKSEKNAIHNLSILEAWSNDPTPYNKDTGKAQLPIDNKLRLSLLYAL 535
QY 522 AKPLNQ-----RSGNPLITNSLVNRTDNDNAETAAPVSPTFRAHDSVQDLADI 573
DB 536 TRPLEKASNKNEIRSLGEPVITNSLNNRSABGKNSERMANYIFTRAHDSVQTVIAKII 595
QY 574 KAEINPNVVGYSFTMEELKKAPEIYNKDLATEKKYTHNTALSYALLLTNKSVPVRYV 633
DB 596 KQAINPKTDGLTFLTDEULKQAPKIYNEWRQAKKKITNSIPTAYALMUSNKSITRLYY 655
QY 634 GDMFTDDGQYMAHKTINYEAEITLLKARIKYVSGQAMRNQOVGNSE-----IITS 684
DB 656 GDMYSDDGQYMATKSPYDAIDTLKARIKYAAGGQDKKIITYEGDKSHMDWDYTGVLTS 715
QY 685 VRYGKALKATDGTDRTRTTSQVAVIEGNNPSRLRKASDRVVVNNGAHKNQAYRPLLT 744
DB 716 VRYGTGANEATDQGEATKTQGMVITSNNPSKLQNQDKVIVNNGAHKNQAYRPLLT 775
```

```
QY 745 TDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVVVPVGAALIKMF 804
DB 776 TKDGLTSTYSDAAAKSLYKRTNDKGLVFDASDIQGLNPNQVSGYLVVVPVGAS--DNQ 833
QY 805 ALRLARPHQOMAS----VHONAAALDSRVMEFGSNFOAFATKKEEYTNVVIKNDVKAPEW 861
DB 834 DVRVAASNKANATGQVYESSALDSQLIYEGFSNFQDFVTQSDYTKKIAQNVOLFWS 893
QY 862 GYTDPEMAPQVYSSYTDGSLDSVIONGYAFTDRYDLGISKPNKYGTGADDLVKAIALHSK 921
DB 894 GVTSEMAPQVYSSSEDSGLDSIIQNGYAFEDRYDLAMSKNNKYSQQDMINAVKALHSK 953
QY 922 GIKVMADWVPDQMYAFPEPEKEVVVTRVVKYGTGPVAGSQIKNTLYVVDGKSSGKQOAKY 981
DB 954 GIQVIADWVPDQIYNLPKEVVVTRVNDYGEYRKDSEIKNTLYAANTKSNKDYQAKY 1013
QY 982 GAFLELOKYPPELFAKQISTGVPMWPSVKIKQNSAKYFNGTNIILGRGAGYVYLKQOATN 1041
DB 1014 GAFLELAAPKPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGYVYLDNASD 1073
QY 1042 TYFNIISDNKEINFLPKTLNODSQVGSYDGGKGVYVYSTGYOAKNTEISEG--DKWY 1100
DB 1074 KYFELKGNQ--TYLPKQMTNKEASTGFVNDGNGMTIFYSTGYOAKNSFVQDAKGNWYFD 1131
QY 1101 NNGYVMTGAQSINGVNYIFLSNGLQRLDAILKNEDGTVAIYNGDGRYRRENGYQF--MSGV 1159
DB 1132 NNGHMYVGLQQLNGEVQVFLSNGVQLRESFLENADGSKNYFGLNRYNGYYSFNDK 1191
QY 1160 WRHFN--NGEMSGLTVIDGQVQVDEMGMGYOAKGKFTTADGKIRYFDKOSGNMYRFRTE 1218
DB 1192 WRYFASGVMAVGLKTINGNTQYFDQDGQYQVKGAWITGSDGKKRYFDGSGNMAVNRFA 1251
QY 1219 NEEGKWLVLGEDAAVTGQTINGOHLVFRANGVOVKGEFVTDHGRISYDGNSGDQIR 1278
DB 1252 DKNGDWYLYNSDGIALVGVQTINGKTYFQDGKQIKGKIITD--NGKLYFLANSSELAR 1310
QY 1279 NRVFNAQCFYFDNNGYAVTGTARTINGQLLYFRANGVQVKGEFVTDYGRISYYDGN 1338
DB 1311 NIFATDSQNNMYFSGDGVAVTGSQTIAGKLYFASDGKQVKGSPVTYN--GKVHYHADS 1369
QY 1339 GDQINRFRVNAQCFYFDNNGYAVTGTARTINGOHLVFRANGVQVKGEFVTDHGRISY 1398
DB 1370 GELQVNRFEADKGNWYLYDSNGEALTGSRINDQVRFTRREGKQVKGVDAYDERLLVY 1429
```

RESULT 10

```
US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRP
; ORGANISM: Streptococcus mutans
US-10-383-930-36
```

```
Query Match      49.3%; Score 3817.5; DB 4; Length 1430;
Best Local Similarity 52.9%; Pred. No. 9e-212;
Matches 762; Conservative 231; Mismatches 394; Indels 53; Gaps 22;
```


Db 180 KYIIGSDGQPKNFPALTNNKVLFDKNTGALTDTIS-QYQFKQGLTKLNDYTPHNVQIV 238
QY 227 DRSAQSEFHVHYLTAEWYRPKYILKOGKTWQSTEKDFRPLMLTWPDQETQROYVNY 286
Db 239 NFENTSLETIDNVYADSWYRPKIDILKNGKTWTASSSEDLRPLLLMSWPDKQTQIAYLNY 298
QY 287 MNAQ-LGINKTYDDTSNQLQNAATAAQIAKIEAKITTLKNTDNLBQTTISAFVKTQSAWN 345
Db 299 MNOQGLGTGENTADSSQBSLAAQTQVQKIEKISQOTQOTQWLRDIIINSEVKTQPNWN 358
QY 346 SDSEKPFDD---DHLQNGAVLYDNEGLTPYANSNYRIILNRTPTNCTGKDKDRYTDADNTI 401
Db 359 SQTESDTSAGEKHLOGGALLYSNSDK-TAYANSDYRLNLRPTSTQTK--PKYFEDNS 415
QY 402 GGYEFLLANDVNSNPVQAEQLNLHFLMNFNIYANDPDANFDSIRYDAVDNVDADLL 461
Db 416 GGYDFLLANDIDNSNPVQAEQLNLHFLMNYGSIIVANDPEANFDGVRVDAVDNVNADLL 475
QY 462 QIAGDYLLKAAKGIHKNDKAANDHLSLILEAWSNDPTYLHDDGDNMINMDNKLRLSLFSL 521
Db 476 QIASDYLLKAHYGVDSKNAIHLHSILEAWSNDPQYKNDTKGAQLPIDNKLRLSLLYAL 535
QY 522 AKPLAQ-----RSGMPLITNSLVNRTDDNAETAAPVPSYFTRAHSEVQDLIADII 573
Db 536 TRPLEKADSNKNEIRSGLEPVITNSLNNRSAGKNSERMANIFIRAHSEVQTVIAKII 595
QY 574 KAEINPNVVGYSFTMBEIKAFEIYKOLLATYKTYHTNTALSVALLLTNKSSVPRVY 633
Db 596 KAIQINPKTDGLFTLDELQAQFKIYNEDMRQAQKYYQTSNIPTAYALMLSNKSITRLYY 655
QY 634 GDMFTDDGQYMAHKIYINYEAITLLKARIKYVSGGQMRNOOVGNSE-----ITS 684
Db 656 GDMYSDDGQYMAKSPYDAIDTLKARIKYAAGGQDMKITIYVEGDKSHMDWDYTGVLTS 715
QY 685 VRYGKALKATDGTTRTTSQVAVLEGNNPSLRKASDRVVNVNCAAHKQOAYPLLLT 744
Db 716 VRYGTANEATDQGEATKQGMVITSNNPSELKQNDKVIYNMGAHKNQEIYRPLLLT 775
QY 745 TNGIKAYSHDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVGVVVPVGAALIKMF 804
Db 776 TKDGLTSYTSDDAAKSLRYKNTDKGELVFDASDIQGYLNPQVSGYLAVVWVPVGS--DNQ 833
QY 805 ALRLARPHQOMAS---VHQNAALDSRVMPGFSNFQAFATKKEEYTNVVVIAKNVDKFAW 861
Db 834 DVRVAASKANATQGVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVQLFSW 893
QY 862 GVTDFEMAPOYVSSYTDGSLDSVIONGYAFTDRYDLGISKPNKYCTADDLVKAIKALHSK 921
Db 894 GVTSEMAPQYVSSYDGGSLDSIIQNGYAFEDRYDLAMSNNKNGYSGQDMINAVKALHS 953
QY 922 GIKVMADVPDQMYAFPEKEVVVATRVKYGTPVAGSQIKNTLYVVDGKSGKQOQAKY 981
Db 954 GQIVADVPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTYKNGKDYQAKY 1013
QY 982 GAFLELOAKYPELPARKQISTGVPMWDSVKIKOMSAKYFNGTNIILRGAGYVLKQDATN 1041
Db 1014 GAFLELAAPYSIENRTQISNGKKIDPSEKITAWKAKYFNGTNIILRGVGYVLKDNASD 1073
QY 1042 TYENISDNKEINFLPKTLNLDQSVQFSYDGGYVYYSYTSYQAKNTFISEG-DKWYFPD 1100
Db 1074 KYFELKQNGO--TYLPQMTNKEASGFGVNDGNGMTFYTSYQAKNSFPVQDAKNWYFPD 1131
QY 1101 NNGYVMTQAQSIINGVNYVYFLSNGQLRDAILKNEIDGTVAAYGNDGRRYENGYYQF-MSGV 1159
Db 1132 NNGHMYVGLQNGEVOYFLSNGVQLRESFLENADGSKNYFGLGNRYSGNYYSFNDISK 1191
QY 1160 WRHEN-NGEMSUGLVTDQVOYFDEMGYQAKGKFTTADGKIRYFDPKQSGMYNRRFIE 1218
Db 1192 WRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKRYFDDGSGNMAVRRFAN 1251
QY 1219 NEEGKWLVLGEDGAATVQSIINGHLYFRANGVQVKGFEVTDHGRISYVDGNSGDQIR 1278
Db 1252 DKGDMWYVNSDGTALVGVTINGKTYFYFGQDGKQIKGKIITD-NGKLKYLANSSEGLAR 1310

QY 1279 NRVFVNAQOMFYFDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDGNS 1338
Db 1311 NIFATDSQNNWYFYFGSDGVAVVTGSQTIACKKLYFASDGQVKGSFVTVN-GKVHYHADS 1369
QY 1339 GQOIRNRFVRNAQOMFYFDNNGYAVTGARTINGHLYFRANGVQVKGFEVTDHGRISY 1398
Db 1370 GBLQVNRFEADKGNWYLDNSGEALTGSORINDORVFFFTREGKQVKGVDYDERLLVY 1429

RESULT 12
US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication NO. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38

Query Match 44.4%; Score 3434.5; DB 4; Length 1554;
Best Local Similarity 46.6%; Pred. No. 1.4e-189;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;

QY 1 MDKVRVYKLRKYKRWVTVSVASAVMTLLTSLGGLVKAD---SNESKSO----- 46
Db 1 MEKLLHYKLVKHYKHWVIAVAS-IGLVSLVGAGTVAEDKVANDTTAQTAVGVDVTDQOQ 59

QY 47 -ISDNTSVV---TANESNVITEATSKQEAASQTNHTVTSSSTSVSNPKEVSVNP 102
Db 60 ATTNDANTNTTDTDTADQSAN-----TNQOQAGSDQSNNOQAKQDTAN----- 103

QY 103 YTVGETASNGEKLQNTT-----TVDKTS---EAAANNISKQTTTADTDVLDSDSNAANL 153
Db 104 -TDRNQADNSQTDNNQATDQATSPATDGTSVQRRDAANVATAADQEGQAPSEQKSAAL 162

QY 154 QILEKLPVVKETIDGKYVYVYDNNKVRNTFTLIADGKILHFD-ETGAYTDTSDTVNKOIV 212
Db 163 ---SLDNVXLIDGKYVYVYVQADGSKYKGFATVNGOMLYFSDTGLALSTSTYSFSG-T 217

QY 213 TTRSNLYKKNQVYDRSAQSEFHVHYLTAEWYRPKYILKOGKTWQSTEKDFRPLMLT 272
Db 218 TNLVDDFSSHNAKYADVSTAKSFELVNGYLTANSWYRPAGILRNGQTWASNENDLRLVMS 277

QY 273 WVPDQETQROYVNYVNAOLGINKT-YDDTSNQLQNLIAAATQAKIEAKITTLKNTDNL 331
Db 278 WVPDKDTQVAVYVNYNKNYLSANETEVNTSETVNETSOVDLNKSAQSQTQKIEKISDNTQWLR 337

QY 332 QTISAFVKTQSAWNSDSEKPFDD---DHLQNGAVLYDNEGLTPYANSNYRIILNRTPTNCTG 389
Db 338 TAMEAFVAAQPKWNSTEN-FNKGDHLOGGALLYTN-SDLTPWANSDYRLNLRPTTQDQ 395

QY 390 KKDPRYTADNTIGGYEFLLANDVNSNPVQAEQLNLHFLMNFNIYANDPDANFDSIR 449
Db 396 TK--KYFTEGEGGYEFLLSNDVNSNPVQAEQLNLHFLMNFNIYANDPDANFDSIR 453

QY 450 VDAVDNVADLLOTAGDYLLKAAKGIHKNDKAANDHLSLILEAWSNDPTYLHDDGDNMIN 509
Db 454 VDAVDNVADLLOVYVNYFDFKQYKVTDSSEANALAHISILEAWSLNDQNTNEDTNGTALSI 513


```
Db 514 DNSSLTSLAVITKPGQRIQLNSLISESVNKRANDTAIGDTIPTYSFVRADHSEVQTV 573
Qy 569 IADIIKAEINPNVVGYSFTMEIEIKAFETYNKDLATEKKYTHYNTALSYALLTNKSV 628
Db 574 IAKIVKEKIDTNSDGYTFLDQLKDAFKLYNEMAKVNTYTHYNI PAAYALLSNMESV 633
Qy 629 PRVYGDMPDQGMMAHTINYEAIETLLKARIKVVGGQAMRNQOV--GNSIETSVRY 687
Db 634 PRVYGDLYTDDQYMAKSPYDAIATMLQGRATVSGGSEEVHKVNGNQNILSSVRY 693
Qy 688 GKGALKATDT--GDRTRTSGVAVIEGNNPBLKASDRVVVNGGAHKNOAYRPLLLTTD 746
Db 694 GDLMSADDTQGTDLSTSGVTLVNSNDPDL--GGDSLTVNMGRAHANOAYRPLILGK 752
Qy 747 NGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGVANPOVSGYLVGVVPGVGAALIKMFAL 806
Db 753 DGVQSYLKDS--TNIVKYTDANGNLTFADDLIGYSTVDMGSLAVWVPVGA--DGDQV 809
Qy 807 RLARPHQQA----SVHQNALDSRVNFEFSGNFQAPATKEEYTNVIAKNVDKFAEMGV 863
Db 810 RVAADTNQKADGKSLKLSAALDSQVIYEGFSNFQDPANNDADYTNKIAENADFFKGLGI 869
Qy 864 TDFEKAPQVSVSTGCSFLDSVLONGVAFTRDYDLGISPKNKYGTADDLVKAIKALHSKI 923
Db 870 TSFEMAPQVVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYSGKDDLALAKALHANGI 929
Qy 924 KMAWDVDPQMAVFEKEVWVTRVDYKGTVPVAGSOIKVTLVVDKSGSKDQQAQYGA 983
Db 930 QAIADWVVDQIYQLPGEEVVTAKTNSYGNPTFFDAYINNALYATNTKSGSDYQAQYGA 989
Qy 984 FLEELQAKYPELPARKQISTGVMPDPSVKIKOWSAKYFNGTNILGRGAGYVLKQOATNTY 1043
Db 990 FLDELKAKYPMFTVMISTGKPIDPSTKIKQWEAKYFNGTNVLCGAGYVLSDDATGY 1049
Qy 1044 FNISDNKEINFLPKLL--NQDSQVGSFDGKGYVYTSYGYOAKNTPISBGDKWYFPNN 1102
Db 1050 FTVNENG--FLPASFTGDQNAKTFYDGTGMAYISTSGNKAVNSFIYEGHYHYFDK 1107
Qy 1103 GVMVTGA--QSLNGVNYFLSLGQLRDAILKNEGTAYAYGNDGRREY--NGYQFM---- 1156
Db 1108 GHWVTSYKAEQNDYFLPNGIQWRDALYQDAQNSYIYGRGTGLYKGDNNWYFPVDPNN 1167
Qy 1157 --SGVWRHFN--NGEMSVGLTVIDGQVYFDEMGOYQAKGFVTTADGKIRYFDKQSGNNMYR 1213
Db 1168 ANKTVRYFDANNVMAIGVRNMYGTYFYFDENGFOAKGQLLTDDKG--THYFDEDNGAMAK 1226
Qy 1214 NRFENESGKLYLGEDGAATGSGTNGOHLIYFR--ANGVQVKGFEVTDHHRISYIDGN 1272
Db 1227 NKV--NVGDDWYMDGNGNAVGQYVNVNNQILYFNPETGVQVKGQFIIDAQGRTSYIDAN 1285
Qy 1273 SGDQIRNRF-----VRNAQGOYFVD--NNGYAVVTGARTINGQL 1309
Db 1286 SGALKSSGFFTPNGSDWYIYAENGYYVKGFKQVAENQDQWYIFDQTTGQAKGAQVGDGR 1345
Qy 1310 LYFRA--NGVQVKGFEVTDYRGRISYIDGNSGDQIRNRVRNAQGOYFDFNNGYAV---- 1364
Db 1346 LYFNPDSGVQVKGDFATDESNTSYFHDGNDGKVVGGFFTCGNNAWYIADNNGNLVKGFG 1405
Qy 1365 -----TGARTINGQHLIYFRA--NGVQVKGFEVTDYRGRISYIDGNSGDQ 1406
Db 1406 BIDGKWHFDEVGTQQAAGALVNGQYLYFVDVSGIQVKGDFVTDGQNTSYIDVNSGDK 1465
Qy 1407 IRNRFVRNAQGOYFDFNNGYAVVTGARTINGQHLIYF--PANGVQVKGFEVTDYRGRISYID 1465
Db 1466 KVGFFTTGDNWYIADQGNLAKGRKSIDNQDLIFDYPATGQVKGQLVSD--GRNYIFD 1524
Qy 1466 ANSGERV 1473
Db 1525 SGSGNMAK 1532
```

RESULT 14

US-10-383-930-40

```
; Sequence 40, Application US/10383930
; Publication NO. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
; US-10-383-930-40
```

Query Match 41.3%; Score 3196.5; DB 4; Length 1518;

Best Local Similarity 45.2%; Pred. No. 98-176;

Matches 695; Conservative 233; Mismatches 472; Indels 139; Gaps 31;

Qy 1 MDKKVRYKLRKVKRWTVSVASAVMTLTTLTSGGL-----VKADSNESKSOISNDSNTS 54

Db 1 MENKHYKLRKVKRWTVSVASAVMTLTTLTSGGL-----VALATVLGGLSVTSSVSADETQKVTYQNSGTT 58

Qy 55 -----VVTAENEENSVITEATSKQEAASSQTNHVTVTSSSSTS 91

Db 59 ASLVTSPEATKEADKRTNTKEADVLTPAKETNAVETATTTNTQATAEAATTATTADVAVA 118

Qy 92 VVNPKEVSVNPTVGETASNGEKLNQNTTVDKTSKTEAAANNISKOTTETADTDVIDDSNA 151

Db 119 AVPNKEAV---VTTDAPAVTTEKAEQPAIV-----KAEVNVTEVKAPEA---ALKDS--- 165

Qy 152 NLQILKLPNVKEIDGKYVYVNNKVRNTFTLIADGKILHFDGTGAVTDTSIDTVNKDI 211

Db 166 EVEAALSANKVKNIDGKYVYVNNKVRNTFTLIADGKILHFDGTGAVTDTSIDTVNKDI 224

Qy 212 VTTRSNLYKYNQVYDRSAQSFHVDHYLTAEWSYRKYILKDGKWTQTOSTEKOPRPLLM 271

Db 225 TTNVVDGFSINRAYDSSEASFELIDGYLTADSWYRPAIIKDGVTWQASTAEDFRPLLM 284

Qy 272 TWPDQETORQYVYVMAQLGINKTYDDTSNQLQNLIAAATIOAKIEAKITTLKQTDWLR 331

Db 285 AWWPNVDQVNYLNTYMSKVFNLDAKYSSTDQETLKVAAKDIOIKIEQIOAKESQWMLR 344

Qy 332 QTISAFVKVTSQAWNSDSEK---PFDDHLONGAVLYDNEGKLTVPVANSYRILNRTPTNQ 387

Db 345 ETISAFVKVTSQAWNSDSEK---PFDDHLONGAVLYDNEGKLTVPVANSYRILNRTPTNQ 403

Qy 388 TG-----KKDPRYTDADNTIGYEFLLANDVNSNPVYQAEQNLHFLMNFNIIYA 438

Db 404 TGTIDKSLDQSDP-----NEMGGFDLLANDVDLSNPVYQAEQNLHFLMNFNIIYA 458

Qy 439 NDPANFDSIRVDAVDNVDADLLQIAGDYLKAAGIHKNDKAANDHLSILSAWSNDPTPY 498

Db 459 GDKDANFDGIRVDAVDNVDADMLQLYTNYFREYGVNKSSEANALAHISVLEAWSLNDHY 518

Qy 499 LHDDGDNINNDKRLSLFSLAKPLNORS--GNPLITNSLVNPTDNDNAET----- 549

Db 519 NDKTDGAALAMENKQRLALLFLSLAKPIKERTPAVSPLYNNTF--NTQDEKTDWINKDGS 577

Qy 550 -----AAVPSYSFTRAHDSEVDLIADIKAEINPNVVGYSFT 587

Db 578 KAYNEDGTVKQSTIGKYNEKYGDASGNVYFIRAHNNVQDIIAEIIEIKKEINPKSGFTIT 637

Qy 588 MEEIKKAFETIYNKOLLATEKKYTHYNTALSYALLTNKSVPRVYVYGMFTDDGQYMAHK 647

Db 638 DAEMKQAFETIYNKOLLATEKKYTHYNTALSYALLTNKSVPRVYVYGMFTDDGQYMAHK 697

Db 758 ANTEGSKYRTSGQVTLVANNPKLNDQSAKLNEMGKIHANOKYRALLIVGTADGIXNF 817
QY 753 HSDQE--AAGLYRTNDRGELIFTAADIKGYANPOVSGYLGWVVPVGA---LIKMFALR 807
Db 818 TSADALIAAGYVKTDSNGLTFGANDIKGETPDMGFFVAVVVPVGASDNQDIRVAPST 877
QY 808 LARPHQOMASVHQNALDSRMFEGPSNFQAF--ATKKEEYTNVVIKQNVDFKFAEWGVTD 865
Db 878 EAKKEGEL-TLKATEAYDSQLIYEGFSNFQTIPIGSDPSVYITNRKIAENVDLFKSWGVT 936
QY 866 FEMAPQVSTDSGLDSVIONGYAFTRDYDLGISKENKYGTADDLVKAIALKALHSKGIK 925
Db 937 FEMAPQVSAADGTFLDSVIONGYAFADRYDLAMSNNKNGYSGKEDRLDALKALHAGIO 996
QY 926 MADWPDQMYAFPEKEVTVATRVDRKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYGCAPL 985
Db 997 IADWPDQIYQLPGKEVTVATRTDAGAKRIADAIIDHSLYVANSKSGKDYQAKYGEFL 1056
QY 986 BELQAKYBELFARKQISTGVPMDDPSVKIKOWSAKYFNGTNILORGAGYVLKQOATNTYFN 1045
Db 1057 AELKAKYPEMFKYNMISTGKPIDDSVKLKQWKALEYFNGTNVLERGVYVLSDEATGKYFT 1116
QY 1046 ISDNKEINFLPKLLNQDSQV-CFSYDGGKGVYVYSTSGYQAKNTFISEGDKWYFDDNGY 1104
Db 1117 VT--KEGNFIPLQLTGKEKVIITGFSDDGKITTYFGTGTQAKSAFVTFNGNTTYFDARGH 1174
QY 1105 MVTGAQ-SINGVNY--FLSNGQLRDLAILKNEDEGTYYAYYCNDRRYENGYQF----- 1155
Db 1175 MVTNSEYSPNGKDVYRFLPNGIMLSNAFYIDANGNTLYNSKGOMYKGGYTKEDVSETDK 1234
QY 1156 ---MSGV--WRHP-NNGEMSVGLTVIDGQVQYFDEMGOYAKGKFVTTADGKIRYFDKQSG 1209
Db 1235 DGKESKVKVFRYFTNEGVMAKGVTVIDGFTQYFEGEDGFOAKDKLV-TFKGKTYFDDAHTG 1293
QY 1210 NMYRNRFIENEKGWLYLGEDGAATGSGTQINGOHLYFRANGVQVKGFEVTDHGRISY 1269
Db 1294 NGIKTW-RNINGKWTYFDANGVAATGAQVINGQKLYFNEGDSQVKGGVVKNADGTYSKY 1352
QY 1270 DGNSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDTRY 1329
Db 1353 KEGFGLVTFNEFTTDDGNVWYAGANGKTVTGAQVINGOHLYFNADGSQVKGGVVKNADG 1412
QY 1330 RYSYDGNSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGOHLYFRANGVQVKGFEV 1389
Db 1413 TYSKYNASTGERLTNEFTTDDGNWYIIGANGKSVTGEVKIGDDTYFFAKDGKQVKGQTV 1472
QY 1390 TDRHGRISYDGNSGDQIRNRFVRNAQGWYFDDNNGYA 1428
Db 1473 SAGNGRISYDGNSGKRAVSTWIEIQPGVYVYFDKNGLA 1511

Search completed: February 11, 2006, 20:56:39
Job time : 140.022 secs

THIS PAGE BLANK (USPTO)

Qy	200	TDTSIDTVN-----KOLVTRSRNLKK-----YNQYDRSRSAQSFEHVDHYLTAE- 243
Db	1034	VSTILDGINLGAATKELLDEHDPLLKKELEAKGVGLAINMNSLSIAATVASIGV--ICAEG 1091
Qy	244	-----SWVRKYILKDGKWTQSTEKDFRPLLMTWVPDQETQRYVNYMN 288
Db	1092	TIFLLPIAGISAGIPSLVNNELIHLDKAT-----SVVNYFN 1127
Qy	289	-----AQLGINKTYDD-----TSNQLQL---NIAAA-----TIOAKIE 318
Db	1128	HLSESKYKGLPKYEDDKILVPIDDLVISEIDFPNNSIKLGTCHILAMEGSGHVTGND 1187
Qy	319	-----AKITVL-----KNWDMLRQTIIS-AFVKTOSAMNSDSEKFPFDHLQN 358
Db	1188	HPFSSPSISGHPSLSIYSIAIGIETENLDFSKKIMLPNAPSVFVFWETGAVPLGRSLEN 1247
Qy	359	GAVLYDNEGKLTIP-----YANSNYRILNRRP-----TNQTKG--KOPR----- 394
Db	1248	DGTRLLDSIRDLVPGKFWRFYAFDFVAILTTLKPVVEDTNIKIKDKOTRNFIMPITTN 1307
Qy	395	-----YTADNTIGGYEFLLANDVNSNPVQAOELNWLHFLMNFIGNIYANDPDANFDS 447
Db	1308	EIRNKLSYFEDGAGTYSLLSSYPSTINILSKDDL-WIF-----1347
Qy	448	IRVDVNDVDADLLQIAGDYKAAKGHHKNDKAANDHLSILEAWSD-----NDTPYLHDD 502
Db	1348	-----NIDNEVREIS-----IENGTIKKGLIKDVLKIDINKNKLIIGNQTI DFSGD 1395
Qy	503	GDN-----MINWMDNKRL-----SLLFSLAKPLNORSQGNPLITSLVNRPTD 544
Db	1396	IDNKORYIFITCELDKXISIIIEINLVAKSYSLLSGDK-----NYLISN--LSNTI 1445
Qy	545	DNAETAAPVPSYFIRAHDSVQDLADIKAEINPNVGVSYFTMBEIKKAFELYN----- 599
Db	1446	EKINTLGDSKNIAINYTDESNNKYFGAISKTQSXIIHYK-----KOSKNILEFYNDSTLE 1502
Qy	600	---KDLIA-----TEKKYTHYNT--ALSAYALLNTKSSV-----PRVY 632
Db	1503	FNSKOFIAEDINVPKMDINTIATGKYVDNNTDKSIDFSISLVSKNQKVNGLYLVNESVY 1562
Qy	633	-----YDGMFTDDQYMAHKTIINYEATIELLLKARIKTVSGGQAMRQN 674
Db	1563	SSYLDFFVKNSDGHHNTSNFNNFLDNISFW--KLFGFENINPVID---KYFT---LVGKT 1614
Qy	675	QVGNSELIITSVRYGKALKATDT--GDRTRTSGVAVIEGNNPSLRLLKASDRVYVNMGA 732
Db	1615	NLGVVERFICDNN-----KNIDIFYGBWKTSKSSKSTIFSGNGRN-----VVV----- 1655
Qy	733	HKNOAYRPLLLTTDNG---IKAYHSDQEAAGLVRYNDRGELIFTAAADI-KGYANPQVS 787
Db	1656	--EPIYNP-----DTGEDISTLSDPSVEPLYGIDRYNK-----VLIAPDLTSLININTN 1704
Qy	788	GYLGVVVPVGAALIKMPALRLARPHQOMASVHQNAALDSRVMEPEGSNFOAFATKKEEYT 847
Db	1705	YYSNEYYP-----EIIVLNPNTFHKVKNINLSDSSPEYKWSATEG-SDFILVRYLBEESNK 1757
Qy	848	NVV-----IAKNVDKFAEWGTYDF---EMAPQYVSSDTGSLDSVIONGYAFTRDYD 896
Db	1758	KILQKTRIKILGNTQSFNKKMSI-DFKDXIKKLSLGYIMSNFKSF-----NSNELDRDH 1810
Qy	897	LGIS-HBNKYGTADDLKAIKALHSKGIKVMADWVPDMYAFV--EKEVYVTAERV---DK 950
Db	1811	LGPKIIDNKYYYDEDSKLVKGLIN-----INNSLYFPDIEPNTVLGHQTINGKK 1861
Qy	951	YGTFPVAGSQIKNTLYVVDGKS-----SGKDOQAQYKGGAFLSE----- 987
Db	1862	YYFDINTGAALTSYKILNGKHGFYFNNDGVNMQLGVPKPGDFGEYFAPANTQNNIEGOAIV 1921
Qy	988	LOAKYPPLFARKQIISTGVMPDPSVKIKQW-----SAKYFNGTNILGRGAGYVLKQDATYT 1042
Db	1922	YQSKFLTLNGKKYYFD-----NNSKAVTGWRIINNEKYFYFNPNAIAA-----V 1965
Qy	1043	YFNISDNKEINFLAPKTLI-----NQDSQVGF-----SYDGKGYVY----- 1077

RESULT 2

US-11-052-554A-217

US 11 021 021 A
; Sequence 217, Application US/11052554A

; Publication No. US20050288866A1

GENERAL INFORMATION:

APPLICANT: Sachdeva, et al.

1. TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
2. TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
3. TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

FILE REFERENCE: 30853/40359A

US/11/052,554A

; CURRENT FILING DATE: 2005-02-05

; PRIOR APPLICATION NUMBER: US 60/589,227

;; PRIOR FILING DATE: 2004-07-20

PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SE

; SOFTWARE: PA

; SEQ ID NO 217

; LENGTH: 726

; TYPE: PRT

Query Match 5.18: Score 395: DB 7: Length 726:

Best Local Similarity 39.3%; Pred. No. 2.9e-15;

Matches	95	Conservative	26	Mismatches	73	Indels	48	Gaps	7
---------	----	--------------	----	------------	----	--------	----	------	---

OV 1214 NRFIENEKGKWL YLGEDGAAVTGSQTINGOHL YFRANGVQVKGEFVTDHHGRISYYDGNS 1273

[illegible]

QY 1274 GDQIRNRFVRNAQGQWFYFDNNGYAVTGARTINGQLLYFRANGVQVKGEFVTDYGRISY 1333

QY 1264 GRISYYDGNQSDQIRNRFVRNAQGWFFYFDNNGYAVTG-----ARTIN----- 1306
Db 2142 -----VQNIIDNNFYIDNNGIVQIGVFDTSQGYKYPAPANTVNDNIY 2183
QY 1307 GQLLYFRANGVQVKGFEV-----TDRY----- 1330
Db 2184 QGAVEY--SGLVRGVEDVYFGETYTIETGWLYDMENESDKYYFVPETKACKGINLIDD 2241
QY 1331 ISYYDGNQSDQIRNRFVRNAQGWFFYFDNNGYAVTGARTINGOHLHYFRANGVQVKGFEV 1390
Db 2242 IKYYPDEKG--IMRTGLISFENNYYFNGEIQGYINIEDKMFYFGEDGVQMIGVENT 2299
QY 1391 DRHGRIYYDGNQSDQIRNRFVRNAQGWFFYFDNNGYAVTGARTING 1437
Db 2300 P-----DGPKYFAHQNTLDENFEGESINYTGMLGLDEKRYFYFTDHYIAATGSVIIDG 2351
QY 1438 QHLYFRANGVQV 1449
Db 2352 BEYFDPDTAQL 2363
RESULT 4
US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-022-562-228

Query Match 3.7%; Score 283.5; DB 7; Length 396;
Best Local Similarity 29.9%; Pred. No. 2.8e-09;
Matches 96; Conservative 44; Mismatches 124; Indels 57; Gaps 17;
QY 1144 DGRYVNGYVQFMGVRHFN--GMS---VGLTVIDQVQYFDEMGOAKGKFTVTTADG 1199
Db 77 DIKLSLGY--IMSN-FKSFENSELDRDLGLFKIIDNKTYTYDEASKLVKG--LININN 131
QY 1200 KIRYEDKQSNRYRNFLENBEGKWLGLG-BDGAATGQSTINGOHLHYFRANGVQVKGFE 1258
Db 132 SLFYFDPDIESLVGT--WQTINGKYYFDINTGAASTSYKIINGKHFFYFNNGVQVKG 189
QY 1259 V-----TDHGRISYYDGNQSDQ-----IRNRFVRNAQGWFFYFDNNGYAVTGARTINGQLLY 1311
Db 190 KPGDGFEPAPANTONNIEGQAIYVQSKFL-TLNGKYYFDNDSKAVTGWTIDGKKY 248
QY 1312 FRANGVQVKGFEVTDYGRISYYDGNQSDQIRNRFVRNAQGWFFYFDNNGYAVTGARTI 1370
Db 249 FNLNTAEATGWQT-----IDGKKYYFNTNTSIATGYTII 284
QY 1371 NGOHLHYFRANGVQVKGFEVTDHGRISYY-----DGNQ--GDQIR--NRFVRNAQGWFFY 1421
Db 285 NGKHFFYFDNNGIMQIGVTF--KGPNGFEYFAPANTDANNIEGQAIYRQNRFL-YLHDNIY 341
QY 1422 FDNNGYAVTGARTINGOHLHY 1442
Db 342 FGNSKAVTGWQTINGNVYF 362

RESULT 5
US-10-873-528-17
; Sequence 17, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-17

Query Match 3.4%; Score 261.5; DB 6; Length 658;
Best Local Similarity 20.7%; Pred. No. 1e-07;
Matches 97; Conservative 72; Mismatches 189; Indels 111; Gaps 18;
QY 1029 RGAGYVLKD--QATN-----TYFNISSNKEINFLPKTLNODSQVGSYDQKGVY 1077
Db 29 QGQYKLEDGSAANWVFDTHYQSWFYI--KADANYAENWELKQ-----GDDYFY 77
QY 1078 YSTSGYQAKNTFISEGDKWYFDNNGYVMTGAQSLNGVNYFSLNGLQRLDAILKNEBGT 1137
Db 78 LKSGGYMAKSEWEDKGAFFYLDQDGKRNWV--GTSYVGTGAKVIEDWVYDSQYDA 135
QY 1138 YAYTGNDGRRYGYQFMGVRHFNNGEMSVGLTVIDQVQYFDEMGOYQAKGKFTVTTA 1197
Db 136 WFIYKADQQAHEKWLQ-----IKGDIYFKSGGYLLTSQWINOA 175
QY 1198 DKIRYFDKQSGNMYRNFLENBEGKWLGLGSDGAATVTSQTI--NGOHLHYFRANGVQVKG 1256
Db 176 ----YVNASGAKVQQGWLFDKQYQSWFYIKENGNAYADKEWIFENGHYIYLLKSGGYMAAN 230
QY 1257 EFVTDHGRISYYDGNQSDQIRNRFVRNAQGWFFYFDNNGYAVTGARTINGQL----- 1309
Db 231 EWIWDKES--WFIYKFDGKMAEKWVYDSSHQAWTYFKSGGY-----MTANEWIWDKESW 283
QY 1310 LYFRANGVQVKGFEVTDYGRISYYDGNQSDQIRNRFVRNAQGWFFYFDNNGYAVTGART 1369
Db 284 FYLKSDDGKIAEKWVYDSSHQAWTYFKSGGYMTANEWIWDKE-SWF----- 328
QY 1370 INGOHLHYFRANGVQVKGFEVTDHGRISYYDGNQSDQIRNRFVRNAQGWFFYFDNNGYAVTGART 1425
Db 329 -----YLLSDGKIAEKWVYDSSHQAWTYFKSGGYMAKNETVDGYQLGSDGKWL----- 377
QY 1426 GYAVTGARTINGOHLHYF-----ANGVQVKGFEVTDYGRISYYDANS 1468
Db 378 -----GGKTTNENAAVQVPTANVYDSDGE-----KUSIYSQGS 413

RESULT 6
US-10-873-528-155
; Sequence 155, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M

```
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 155
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-155

Query Match      3.4%; Score 261.5; DB 6; Length 677;
Best Local Similarity 20.7%; Pred. No. 1.1e-07;
Matches 97; Conservative 72; Mismatches 188; Indels 111; Gaps 18;

QY 1029 RGAGVYVKD--QATN-----TYFNISDNKEINFLPKTLINQDSQVGFSDGKGVY 1077
Db 48 QKQYLKEDGSQANEFVDTHTYQSWFYI--KADANYAENELKQ-----GDYFY 96

QY 1078 YSTSGYQAQNTTISEGDKWYFDNNGYMTGAQSGINGVNYFLSNGLQLRDAILKNEGT 1137
Db 97 LKSGGYMAKSEVEDKGFYLDQDGKMKRANV--GTSYVGATGAKVIEDWYDSQYDA 154

QY 1138 YAYYGNDRRYENGYQFMGVRHFNNGEMSVGLTVIDGVOYQFDEMGOAKGPFVTTA 1197
Db 155 WFYIKADGQHAKEWLIQ-----IKGXDYFKSGGYLLTSQWINQA 194

QY 1198 DGKIRYFDKQSNWYRNRIENECKWLYLGEDGAVTGSQTI--NCHLYFRANGVQVKG 1256
Db 195 -----YVNSAGAKVQOGWLFQYQSWFYIKENGNYADKEWTFENGHYLYLKS GGYMAAN 249

QY 1257 EFVTHHGRISYDNGSGDQIRNRFVRNAQGO--WFYFDNNGYAVTGARTINGOL----- 1309
Db 250 EWIWDKES--WFLKFDGMAKEWYDHSQAWYFKSGGY-----MTANEWIWDKESW 302

QY 1310 LYFRANGVQKGEFTVDRGRISYDNGSGDQIRNRFVRNAQGOQWYFDNNGYAVTGART 1369
Db 303 FYLKSPGKIAKEWYDHSQAWYFKSGGYMTANEWIWDKE--SNP----- 347

QY 1370 INGCHLYFRANGVQKGEFTVDRGRISYDNGSGDQIRNRFVRNAQ-----QWYFDNN 1425
Db 348 -----YLKSDGKIAKEWYDHSQAWYFKSGGYMAKNETVDGYQLGSDGKWL----- 396

QY 1426 GYAVTGARTINGOHLFYR-----ANGVQVKGEFTVDRYGRISYDANS 1468
Db 397 -----GGKTNENAAYYQVVPYANVYDSGE-----KLSYISQGS 432

RESULT 7
US-11-013-759-11
; Sequence 11, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 11
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-11

Query Match      3.2%; Score 251; DB 7; Length 2314;
Best Local Similarity 19.4%; Pred. No. 2.4e-06;
Matches 346; Conservative 182; Mismatches 655; Indels 596; Gaps 79;

QY 38 ADSNESKSOISNDS-----NTSVVTANEESNVITEATSKQEAAS-----SQTNHTVTTTS 86
Db 714 ASGNDTKNIIRGLSPPLPSITNAGGVRTTEQGNITISDEDEKSKAAASIGILNTPGNLKN 773

QY 87 SSSTSVNPKVSVNPNYTVGETASNGEKLNQNTTV-----DKTSEAAANNISQJTTE 139
Db 774 SNSVGFVSTYNTVDFIDGNATTAKVYDETQTSKYTYDVNVDEKTIETLTDGNGKTNKIG 833

QY 140 AUTDVIDDSNAANLQILEKLPNVKEIDGKYVYDNNQKVRNFTLIADGKILHFDGTGAY 199
Db 834 VKTTLTITTTNA-----NGKA--TNFSTTDNDALVNAKDIA-- 866

QY 200 TDTSIDTVNKDITVTRSNLYKKYNQVYDPSAQSEFHVHDHLYLTAESWYRKYILKDGKTWT 259
Db 867 --ENLNTLAKEIHTKGTGTA-----DTALQTEK-----VKDQGATDD 900

QY 260 QSTEKDFRPLMTWPFQDQETQRYVNYM-----NAQLGINKTY 297
Db 901 ET-----ITVGKDGTONKGTNTVTLKKGNGLTVATNKDGTVTFGINTQSGLKAGD 951

QY 298 DPTSQNLQINIAAATQAKIE-----AKITTLKNTDMLRQTISAFVKTQSAMNS--- 346
Db 952 STTLNKGDLGSIKNPASNEQIQVGADGVKFAKVGKNSGSTIGDT--SRITKQDQIGFTGANG 1010

QY 347 --DSKPP--DDHLQNGAVLYDNEG-----KLTPY-----ANSYRILN----- 381
Db 1011 SLDTTKPHLTCKLKVGEVEITNTGINAGKKITNIQSGDITQNSNDVATGGRVYDLKTE 1070

QY 382 -RTPNTQCKPKDPRVTADNTIGGYEFLLANDVDSNPVQAEQLNWLHFLMNFNIYAND 440
Db 1071 LESKINSAAK-----TAQNSL--HEFSVADEQGN-----HFTV-----SN 1103

QY 441 PDANFDSIRVDADVNDVADLQIAGD---YLKAAKGIHK--NDKAANDHLSILEAWSND 495
Db 1104 PYSSYDTSKT-----SDVITFAGENGITTKVNGKGVVRVGIDQTKGLTTPKLTGVNNGN 1156

QY 496 TPLYLD--DGDNMNMDNKLRLSLFLSLAKPLNQSGMPLITNSLVNPTDNN--AETAA 551
Db 1157 KGIVIDSKDQNTIIT-----GLSNTLANVTN--DGAGHALSQGLANDTDKTRAASIGD 1207

QY 552 VPSYFIRAHDSVEQDLIA-----DIKAEINPNVVGYSFTMEEIKKAFEIY--NKDLLA 604
Db 1208 VLNAGFNLOQNGEAVDFVSTYDVFIDGNATTAKVYDVTDSKTSKVYDVNVNDKNTIEV 1267

QY 605 TEKK-----YTHYNT----- 614
Db 1268 TSDKGLGVKTTTLTKTSANGNATKFSAADGDALVKASDIATHLNTLAGDIQTAKGASQAS 1327

QY 615 -ALSVALLLTNK---SSVPRVYVGMFTDDGGYMAHKTINYEAI-----ETLLKARI 662
Db 1328 SSASYVDADGNKVIYDSTDKKY--QVNDKGQVDKNKEVAKDKLVAQAQTPDGTTLAQMNV 1385

QY 663 KYVSGQAMRNQOVGNSEIITSVRYGKGALKATDGTDRTRTSGVAV----- 709
Db 1386 KSV-----INKEQVNDANKKQGINEDNFAFIKGLENAAKDQTKTNAAVTVGDLNVAQTP 1440

QY 710 -IEGNNPSLRKASDRVVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYT--- 765
Db 1441 TFGADGTGTTAKKGLGTLTIKGGQDTNK-----LTDNNIGV-----VAGTDGFTVKL 1487

QY 766 -----NDRGELIIFTAADIKGYANPQVSGYLGWVPVGAALIKMFALLRLARPHQOM 815
Db 1488 AKDLTNLSNVNAGG---TRIDEKGISFVDANGQAKANTPVLISA---NGLDLOGKRIISNIG 1541
```

```
QY 816 ASVHQAALDSRVMEGFSNFOAFATKKEEYTNVVIKAVDKFAEWGVTDPFEMAPQYVSS 875
Db 1542 AAVDDNDV-----NFKQFNEVAKTVNNLNQSNQSGASLFPFVTTDANGKP--ING 1589
QY 876 TDGSLDSVIQNGYAFTRDYDLIGISKPNKYGTADDLVKAIKALHSGIKVMADWPQDMY 935
Db 1590 TDG-----KPO-----KAIKAGDKYHANANGVPVVDKD 1618
QY 936 APPEKEVVTATRVKYGTPV-AGSQ-----IKNTLYVVDGKSSGKQQA 978
Db 1619 GKPIITADKLANLAHAGKPLDAGHGVVASLGSGSDAITLTNIKSTLPQIDTPNTG--NA 1675
QY 979 KYGGAFLLEQAKYFELFARKQIS-----T 1003
Db 1676 NAGQA-----QSLPSLSAAQCSNAASVKDVLNVGNLQTNHNVQDFVKAYDVTNFMVNGT 1729
QY 1004 GVPM-----DPSVKIKOWSAKYFNGTNIILGRG--AGYVLK 1036
Db 1730 GADITSVSADGTMSTNITVNTALAATDDGNNVLIKAKGKFYKADDLMPNGLKAGKAS 1789
QY 1037 DQATNTYFNI-----SDNKEINFPLKTLNQD-----SQVGFSYDGGK--GYV 1076
Db 1790 DAKTPTGLSLVNPAGKSGTGDAVALNLSKAVPKSKDGTTTTTSVSDGISIQGKDNSSI 1849
QY 1077 YSTSQYQAKNTFISEGDKWYFDNNGYMTGAQSGINGVNYFLSNGQLRDAILKNEDEG 1136
Db 1850 TLSKDKQLNVGKVI SNVGK-----GTRKTDAAANVOQLNE--VRNLLGLGNAGNDNADG 1900
QY 1137 TVAYY-----GNDGRYENGYQF-----MS 1157
Db 1901 NQVNADIKDPNSGSSNRVTIKAGTVLGGKNNDEKLAGGVQGVQVGDVGDGNANGDLS 1960
QY 1158 GVM-RHFNNGEMSVGLTVID--GVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMY-- 1212
Db 1961 NVWVKTKDGGSKALLATYNAAGQTNLTNNPAEADRI--NEQIRFEHVNDGNQEPV 2017
QY 1213 ---NRFTENEKGK-LVLG-----BDG--AAVTGSQTINGHLYFRANGVQVKG----- 1257
Db 2018 VQGRNGIDSSAGKHSVAIGFQAKADGEAAVAIGRQTOAGNQSAIGDQAATGDSIAI 2077
QY 1258 ---FVT--DHRGRI-----SYDNGSGDQIRNRFVRNAQGWYFVFNNDN----- 1295
Db 2078 GTGNVVTGKSGAIGDPTVTKADNSYSVGN-----NQFIDATQDVFVGNNITVTESN 2132
QY 1296 -----GYAVTGARTINGQLLYFRANGV--QVKGFEVTDYGRISYYDNGSGD 1340
Db 2133 SVALGSNSAISAGTHAGTQAKKSDGTAGTTTAGATGTVKGFAGQTAAGVAVSV--GASGA 2190
QY 1341 QIRNRFVRN-AGQGWYFDDNNGYAVTGARTINGHLYFRANGVQVKGFEVTDHGRISYY 1399
Db 2191 E---RRIQNVAAE-----VSATSTDVANGSOLYKATQGI---ANATWELDHRIHQN 2236
QY 1400 DNGSGDQIRNRFVRNAQGWYFDDNNGYAVTGARTINGQ 1438
Db 2237 ENKANAGISSAWAMASMPQ-AYIPGRSVMTGGIATHNGQ 2274

RESULT 8
US-11-052-554A-229
; Sequence 229, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 91
; LENGTH: 2902
; TYPE: PRT
; ORGANISM: Helicobacter pylori J99
US-11-052-554A-91
Query Match 3.2%; Score 247; DB 7; Length 2902;
Best Local Similarity 17.8%; Pred. No. 5.6e-06;
Matches 367; Conservative 254; Mismatches 646; Indels 794; Gaps 96;

QY 30 TLGSLVKADNENKESQISNDNTSVVTANEESNVITEATSKQEAASQTNHVTTS88S 89
Db 439 TISGVTLSGKNDLKGATLDFGSSKIT-----LTQGT-----FNLTSIGSEKS 483
QY 90 TSVNPKVEVSNPVTYVGTASNGEKLQNTTVDKTSAAANNIS-----KQTEADT 142
Db 484 VTILNSRGGITYNHLLNHAI-----NSLTNALKTNESSSKPSQFAOGLMDMTYNGVT 536
QY 143 DVIDDSNAANIQILEKLPNVKEIDKYYYY-----DNNKVRTNFTLIADGKILHFDFT 196
```

```
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 229
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-229
Query Match 3.2%; Score 248; DB 7; Length 619;
Best Local Similarity 24.3%; Pred. No. 5.5e-07;
Matches 84; Conservative 35; Mismatches 129; Indels 98; Gaps 13;

QY 1056 PRTLLNDPSOVGFSYDGKGYVYVTSYGQ-----AKNTFISGDKWYFDNNGY 1104
Db 360 PKPEKTDQQAEEDYARRSEBEYNRLTQQPPKAEKPAPAKTGKQENGMMYFYNTDGS 419
QY 1105 MVTGAQSGINGVNYFLSNGQLRDAILKNEDEGTVAYVYNGDGRRRYENGNGYVYVTSYGQ 1164
Db 420 MATGWLQNNQSWYILNSNGAM-----ATGWLQY-NGSWYILN 455
QY 1165 -NGEMSVGLTVIDGVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNRPIENEKG 1223
Db 456 ANGAMATGWAKVNGSWYILNANG-----AMATGWLQY-----NGS 490
QY 1224 WLYLGEDGAAVTGSQTINGHLYFRANGVQVKGFEVTDHGRISYYDNGSGDQIRNRFVR 1283
Db 491 WYILNANGAMATGWAKVNGSWYILNANGAMAT-----GWLQY----- 527
QY 1284 NAQGWYFDDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDNGSGDQIR 1343
Db 528 --NGSWYILNANGAMATGWAKVNGSWYILNANGAMATG-WVKD--GDTWYILEASGAMKA 582
QY 1344 NRVNNAQGWYFDDNNGYAVTGARTINGHLYFRANGVQVKGFEV 1389
Db 583 SQWFK-VSDKWWYVYVNGLG-ALAVNTTVDGYK-----VNANGEWV 619
```

RESULT 9

```
US-11-052-554A-91
; Sequence 91, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 91
; LENGTH: 2902
; TYPE: PRT
; ORGANISM: Helicobacter pylori J99
US-11-052-554A-91
```

```
Query Match 3.2%; Score 247; DB 7; Length 2902;
Best Local Similarity 17.8%; Pred. No. 5.6e-06;
Matches 367; Conservative 254; Mismatches 646; Indels 794; Gaps 96;

QY 30 TLGSLVKADNENKESQISNDNTSVVTANEESNVITEATSKQEAASQTNHVTTS88S 89
Db 439 TISGVTLSGKNDLKGATLDFGSSKIT-----LTQGT-----FNLTSIGSEKS 483
QY 90 TSVNPKVEVSNPVTYVGTASNGEKLQNTTVDKTSAAANNIS-----KQTEADT 142
Db 484 VTILNSRGGITYNHLLNHAI-----NSLTNALKTNESSSKPSQFAOGLMDMTYNGVT 536
QY 143 DVIDDSNAANIQILEKLPNVKEIDKYYYY-----DNNKVRTNFTLIADGKILHFDFT 196
```


Db 537 GOLLNENATSPKTSSTNSTQVYQVGIKIGDTIYKQETPS--HNSIIIOALES 594
QY 197 GAYTDTSDIVNKDIIVTTRSNLYKYNQVYDRSAQSFEHVDHLYTAESWYRPKYILKCKG 256
Db 595 GYTPPPVINGSK-----FDLSASNYINAD-----MPWYHXYIPKQ 633
QY 257 TWQSTKDFRLLMTWPDQBTORQYVYMYNAQL--GINKTYDDTSNOLQNLIAAATIQ 314
Db 634 NPTESCTY--YLPVSQIWGSYTNFSKQTFASNSNLVIGYNATWTDH-----NVSSSDTV 686
QY 315 AKIEAKITTLKN-----TDWLRQTIISAF-----VKTQSAWNSDEXP 351
Db 687 AFGDTSGSALNGHCGWPYPYQCTGTNGTYSAHYVYITANLSRGNRIGTGGAANLIFNGV 746
QY 352 FPDHLQNGAVLYDNECKLTPY-----ANSNYRIL--NRPPTNOTGKDP 393
Db 747 DSINATNATITOHNAGAYSSMTFTQNMNDNSQNLNGLNSKLLVYGTFTTQNA--KDG 804
QY 394 RY-----TADNTI--GGYEFLANDVNSNPVVQABQLNWLHFLMNFNIYANDPDA 443
Db 805 KEIFNAGQATFENTFNNGSYQF--SGDSLNFN-----NNQFNSGGFEIGAKTIFN--NA 857
QY 444 NFDSIRVDADVNDLQIAGDYLKAAKG--IHKNDKAAHDLSLLEAWSNDPTPLHD 501
Db 858 NFNNSSFSNFNSSA--TTSFVGDTNANSNLQIAGNAVFNGSTNGSQNTANFNNTGSVNI 916
QY 502 DG-----DNMI-----NMDNKLRLSL--LFLSLAKPLNORSQWNPILITNSLVN-----RTD 544
Db 917 AGNATFDVNFVNSPNTSVKGVYLLNITLKNLAPLSPFGDTIVFSAHSVINIGEAITN 976
QY 545 DNAETAAPVSYSF-----IRAHDSEVQDLIA-----DII-----573
Db 977 GNPIITLVSSSKAIEYNDAPSKNLWOLINYQHGASSEKLVSSAGNGVYDVVYSFNQVTN 1036
QY 574 -KAEINPNV-----VGSFTWEEIKKAFEYINKDOLLATEKKYTHYNTALYALLTNK 625
Db 1037 FOEVFSPNSISIRRLGVGVYFVYDMEKSDRLY-----YQNALGFMTYMPNS 1083
QY 626 -----SSPRVYVYDMDTDCQYMAHKTINYEATETLLKARIKYVSGQAMRNOQV--G 677
Db 1084 YNNNLGNLNTIYYDNSID--FYASGKTL-----FTKAEFSQTFQ-----NSAIVPG 1131
QY 678 NSEIITSV-----RYG--KGALKATDGTDRITRTSGVAVIEGNPNLSR-----L 719
Db 1132 AKNIWTSVSDAPQSNVIIRFGDNKGA--GSDASGHCWNLQCIGFITGHVEAQKIYITGSI 1190
QY 720 KASDRVVMGAAHKNQAVRPLLLTDDNGIKAYHSDQEAAGL-----VRYTNDRGEI 771
Db 1191 ESGNRISGGGASLNFNGLGILLT--NATLYN--RAAGTQSSSMNFVNSANIOAQN 1244
QY 772 ---IFTAADIKG-----YANPQVSGYLG-----791
Db 1245 SYFIDDTAQNKNPNFNSFALMLDPSNSFRYQGTQSVFKFNVAIVNSFTNSSNLSSG 1304
QY 792 -----VWVPVGAALIKMFALRLARP-----811
Db 1305 LYQMOAKSVLFDNSNLVSQVGTSSIKANAINLSQNASINASHSTLEQGLDLNLDTSGL 1364
QY 812 ---HOQMASVHQAALD-----SRVMEGFSPN--QAPATKKEEYTNVVIKKNVDPK 858
Db 1365 NLNQSAIVNSNATINDYASLIASNGSHLNFNGAVNFNSANITTSLSSSSIIVFKGAVSLR 1424
QY 859 AEWGVT-----DFEMAPQVVSSTDCSFLDSDVI--QNGYAFTRDYDL-----GLSKPNK 904
Db 1425 GQFNLSNNSLDFQSSAITSNATFNFDNAPFSQPIFTHQALDIKVPFLSGNLNLNPN 1484
QY 905 YGTAD-----DLVKAIK--ALHSKGIKVMADWPDQWYAPPEKEVTA-----TRVDKYG 952
Db 1485 SSVLNLKNSQLVFSQGSINIANIDLLSDLNGKNRVY--NIQADMNGWYERINEFG 1541
QY 953 TPVAGSQIKNTLYVVDGSSGKQQAQYGGAFLEELQAKYFELFARKQISTGVPMDPVK 1012

Db 1542 -----MRINDIYDAKNQYVSFTNPLNNAL--KITESFKNNQLSVTLISQIPGK 1588
QY 1013 IKQSAKYFNGTNIILGRGAGYVLKQATNTYFNISDNKEINFLPKTLNQDSQVGSYDG 1072
Db 1589 -----NTLYNI--GSEIFNYQKYNANGVYSDDDA 1618
QY 1073 KGYVYVYSTGYOAKNTFISEGDKWYFDNNGVMVTGAQSGINGVYVYFLSGLQLRLDAILK 1132
Db 1619 QG--VFVLTSSV-----KGYNPNQSQOASGS-----NNTTKNNLTSESSVIS 1660
QY 1133 NEDGTIYAYGN--DGBRYENGYQPM-----1156
Db 1661 Q--TINAQGNFISALHVTNKGYNFSNIKALQOMALKYPEIKKILGDNFSLSSLNKLG 1717
QY 1157 -----SGVRRFPNNGEMSVGLTVIDQVQV-----YFDEM 1185
Db 1718 DALNQLTKLITPSDMKNINELIDNANNSVQVFNNGTLIGATKI--GQTDNLSAVVFGGL 1776
QY 1186 GYOAK-----GKFTVTADKIRYFDKQ-----SGNMY-----1212
Db 1777 GYQKPCDVTDIVCQKFRGTGLQLLESISADLGYIDTTFNAKEIYLTGTLGSGNMGTCG 1836
QY 1213 -----RNRFIENE-----EGKWLILGEDG-----AAVTGS--QTINGQ 1243
Db 1837 SASVTNFSQTSLLNQANIVSSQTDGIFSMLGQEGINKVFNQAGLANILGEVAMQOSIN--1894
QY 1244 HLYFRANGVQVKGFEVTDHHRISYYDGNSGDQIRNRFVRNAQOGWIFYFDN-----1294
Db 1895 ---KAGGL--GNLIVNLTGSDSVIGVLTPEQKNQTLSQLLQ--NPFNMLMNDSLMT 1946
QY 1295 -----NGYAVTGAFTIN-----1306
Db 1947 AIKOLIRQKLGFWTGLVGLAGLGGIDLQNPKEKLIGSMISINDLLSKKGLFNQITGFISAN 2006
QY 1307 --GOLLY-----FRANGVQVKGEP-----VTDY 1328
Db 2007 DIGQVISVLMQDIVKPSDALKNDVAALGKQMTGEFLQDPTLSLESLLQNOQIKSYLDKV 2066
QY 1329 ---GRISYDGNSDQIRN-----GVQV-----KGEFVTDHR 1393
Db 2067 LAAGLGLSIEGGLGDLIPNLKKGIFAPYGLSQVWQKGFDFNAQGNVFNQNSFTSNAN 2126
QY 1361 GYAVTGAFTIN--GOHLYFRAN-----GVQV-----KGEFVTDHR 1393
Db 2127 G--GTLFSNAGNSLIFAGNHIAFTNHSGLTNLLNSQVSNINVTMLNASGLKINATN 2182
QY 1394 GRISYDGN-----SGDQIRNR-----FVR 1413
Db 2183 NNVSYSQGNLFINASCVOQSDPTTASATNPCTTAQNASSSNASNAPIALNNDESIVV 2242
QY 1414 NAQOGWIFYDNNGYA-----VTGARTINGOHLY-----FRANGVQVKG-----F 1453
Db 2243 TANG--FNFSNIIYANGVDFSKIKGSANVKULIYNNNAQFOANNLTITSNQAVLEKNASF 2300
QY 1454 VTDYGRISYDYDANSERVRI 1474
Db 2301 VTNNLNIQGAFFNNATQKIEV 2321

RESULT 10

US-11-052-554A-231
; Sequence 231, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004


```

; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 228
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-228

Query Match      2.9%; Score 221; DB 7; Length 338;
Best Local Similarity 21.9%; Pred. No. 8.2e-06;
Matches 84; Conservative 59; Mismatches 130; Indels 110; Gaps 18;

QY 1117 YFSLNGLQRLDAILKNEGDGTYAYYGNDRGRRYENGYYQFMSGVWRHFNNGENSVGLTVID 1176
Db 15 FFF---GLLATNTVANTTGG-RFVDKDKRKY-----YVKDDHKAIIYWHKID 57

QY 1177 GQVQYFDEMGYQAKGKFTTADGKIRYFDKQSGNNYRNFINE-----BKGWLYLGE 1229
Db 58 GKTYFFGDIGEMVVG-----WQYLE-IPGTGYRDLFDNPQVPVNEIGLQEKWYFQG 107

QY 1230 DGAAY--TGSQTFING-----OHLIFRANGVQVKGFEVTDHGRISY 1268
Db 108 DGALLEQTDKQVLEAKTSNTGKVGEOYPLSAEKRTTYFDNNYAVKTGWIYE-DGNWY 166

QY 1269 YD--GNSGDOIRNRFV-----RNAQOGWFYDNNGYAVTGARTING 1307
Db 167 LNKLGNGFDDSDYNPLPIGEVAKGTQDFHTVIDIRSKPAPWYLDASGKMLTDWKVNG 226

QY 1308 QLLYFRANGVQVKGFEVTDYGRISYDNGSDQIRNFRVNAQOGWFYFDN--NGYAVTG 1366
Db 227 KWYFFGSSGSMATG-----XKVR--GKWWYLDNKGDMKTG 261

QY 1367 ARTINGOHLIFRANGVQVKGFEVTDYGRISYDNGSDQIRNFRVNAQOGWFYFDN 1425
Db 262 WQYLGKWKYLLRASSGAMVTGWY---QDGLTWYLLNAGNDMKTGWQVFN--GKWWYAYS 316

QY 1426 GYAVTGARTINGOHLIFRANGVQ 1448
Db 317 G-ALAVNTTVDGYSVNYNGEWVQ 338

RESULT 14
US-11-013-759-3
; Sequence 3, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:Jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-3

Query Match      2.8%; Score 219; DB 7; Length 1992;
Best Local Similarity 20.0%; Pred. No. 0.00013;
Matches 351; Conservative 215; Mismatches 665; Indels 524; Gaps 93;

QY 30 TLGGGLVKADS--NESKQISQSDNNTSVVTANEESNV-----TEATSK 71
Db 119 TLINDLINGHVLKEIRSSKNDVKYRRTTAGSHASTAGVMSYAQGHFSNAGFGRATATK 178
```

```

; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 228
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-228

Query Match      2.9%; Score 221.5; DB 6; Length 744;
Best Local Similarity 29.1%; Pred. No. 2.3e-05;
Matches 69; Conservative 29; Mismatches 118; Indels 21; Gaps 10;

QY 1218 ENEBKGWLYGBDGAATVGSQTINGOHLIFRANGVQVKGFEVTDHGRISYDNGSDQI 1277
Db 529 KGENGWYFYNTDGSMAICWLNQNGSWYLLNANGAMATG-WVKD--GDTWYLEASGAMK 585

QY 1278 RNRFRVNAQOGWFYDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYDGN 1337
Db 586 ASQWFK-VSDKWKYVYNSNGAMATGMLQYNGSWYLLNANGDMATGWL--QYNGSWYLLNA 641

QY 1338 SGDOIRNFRVNAQOGWFYDNNGYAVTGARTINGOHLIFRANGVQVKGFEVTDHGRIS 1397
Db 642 NGDMATG--WAKVNGSWYLLNANGAMATGWAKVNGSWYLLNANGSWATG-WVKD--GDTW 696

QY 1398 YYDNGSDQIRNFRVNAQOGWFYDNNGYAVTGARTINGOHLIFRANGVQVKGFEV 1454
Db 697 YLEASGAMKASQWFK-VSDKWKYVYNGLG-ALAVNTTVDGK-----VNANGEWV 744

RESULT 13
US-11-052-554A-228
; Sequence 228, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
```

QY 72 QE-----AASSQTHVTSSST-----SVNPKV 98
DB 179 SAYSLAVGLAATAEGQSTIAIGSDATSSGLAIALGAGTRAQLOGSIALQGQSVVTSQDN 238
QY 99 VSNPVTGETASGKQLQNTTVDKTSSEAAANNISKO--TTEADTVDDSDNAANLQIL 156
DB 239 NSRPAYTNTQALDPKFAQNTWKAGPLSIGNSIKRKIIINVAGVKNKTDAVNAQLEAV 298
QY 157 EKLPNVKEIDGKYVYDNG-----KVRTNFTLIADGKILHFDGTAYTSDTIVNK-DIV 212
DB 299 VKWAKERRIT---FQDDNDSTDKGLDNLTLTKGA---ETNALTDNNIGVVKREADNS 351
QY 213 TTRSILYKYNQVDRSAQSFHVDHLYTASWYRPKYLKDGKTWTQSTEDKFRPLMT 272
DB 352 GLKVKLAKTLNLTENVNTTLNATTVKVGSSSTTAEALLSLSLTFTQFNT-----402
QY 273 WMPDETORQY-----VNYMN---AOLGINKTYDD-----TSNQL 304
DB 403 ---GSQSTKTVGVNGVKFTNNAETTAALGTTRITRDKIGFARDGDVDEKQAPYLDKKQL 460
QY 305 QLNIAAATTOAKIEAKITLKNVDLWROTISAFVKTQSAWNS---DSEKPPDDHLQNGAV 361
DB 461 KVGSAIITIDNGIDAG-----NKKISNLAKGSSANDAVTIEQLKAAKPTLNAGAG 510
QY 362 LYDNEGKLTTPYANSYRILNRT-PTNQTGKKDPRYTADNTIGGYEFLLANDVDNSNPVQ 420
DB 511 ISVTPEISVDAKSG---NVTAPYNTIGVKTTLSNDSGTSKFS---VKGSGTNSLVT 563
QY 421 AEOL-NWLH-----FLMNFIGNIYANDPDANFDSIRVADVNDVA-DLLQIAG-----465
DB 564 AEHLASYLNEVNRADALSQSFVKEEDDDANALTVAKDITKNAGAVSILKKGKGLT 623
QY 466 -----DYLKAAKGTHKND-----KAANDPHLSILEAWSNDTPLYLHDDGN 505
DB 624 VATKKGTVTFGLSQDGLTICKSTLNDGLTVKDTNEQIQV---GANGIKFTNVNGSN 679
QY 506 ---MINMDNKLRLSLLFS-----LAKPLNQRSGM---NPLIINSLVNRTDNDNAETAA 551
DB 680 PGTGIANTARITRDKIGFAGSDGAVDTNKPYLDDQKLGQVGNVKITNTGIN-AGGRAITGL 738
QY 552 VPSYSFIRAHDS---EVODLIADIHK---AEINPNVGVSTFMBEIKKAFE-IYNKDL- 603
DB 739 SPTLSIADQSGRNLIELGNTQDKDKSNAASIN-DILNTGFLKNNNRPIDFVSTYDIDV 797
QY 604 -----ATEKKYTHYNALSALLLNKSSVPRVYGDMTDDGQYMAHKTINYEAITLL 658
DB 798 FANGNATTATVTH-DTA-----NKTS---KVVY-DVNVDD-----TTIHLTGTDNKK 839
QY 659 KARIKYV-----SGQAMRNOQVGNSE-----IITSVRYGKGALKATD 696
DB 840 KLGVRTTKLNTKSANGNTATPNVNSSDEDALVNAKDIAENLNTLAKSIHTTKG-----893
QY 697 TGDRTTRTSVGAVIEGNNPSRLKASDRVVMGAAHKNQAVRPLLLTTDNGIKAYHSDQ 756
DB 894 TADTALQFTVKKVDENI-----NADDANAIVTGQKNANNQVNTLTKGENGLNI-----943
QY 757 EAAGLVRYTNRGELIF---TAADIKYANPOVSGYLGWVVPVGAALIKMFALRLARPHQ 813
DB 944 -----KTDKNGTVPFGINTTSLGKAGKSTLNDGGLSIKNPTGSEQIQVGA-----DGV 991
QY 814 QMASVHQN-----AALD--SRWFE---GFSNFOAFATKKEEY-----TNV---849
DB 992 KPAKYNNGVGVAGIDGTTRIDREIGFTGTNGSLDKSKPHLSKDGINAGGKKITNIQSG 1051
QY 850 VIAXN-----VDKFAEWGVTDFEMAPOY-----VSSTD 877
DB 1052 EIAQNSHDAVGGKLYDLKTELENKISSTAKTAQNSLHFEVSADQGNFTVSNPSSYD 1111
QY 878 GSFDSVI---QNGYAFTRDYDLGISKPNKYGTADDLVKAICA-----LHSGKIKV- 925
DB 1112 TSKTSDVITFAGENG--ITTKVNGKGVW----VGIDQTKGLTTPKLTVGNNNGKGIVID 1164

QY 926 -----MADWVPDMYAPPEKEVVVTVTRVDKYGTPPVAGSQIKNTLYVVDGKS 971
DB 1165 SONGNTITGLSNTLANVTNDK-----GSRVTEQGNIIKDEKTRAASIVDVL 1214
QY 972 SGKDOQ-----AKYGGAFLEELQAKYPFARKQISTGVPMDPDS---VKIK 1014
DB 1215 AGFNLQNGEAVDFVSTYDVTNFPADGNATTAKVTVDDTSKTSKVVDVNVDDTTIEVKDK 1274
QY 1015 QMSAKYFNGTNI LRGAG---YVLKQAT-NTYFNISD-NKEINFL-----PKTLLNQD 1063
DB 1275 KLGVK---TTTTLTGTGANKFALSNQATDALVKASDIVAHLNLTSGDIQAKGASQAN 1331
QY 1064 SQVGF-SYDGKGYVYVSTSG---YQAKNTFFISEGDKWYFDDNGYVMTGAQSLNG---VN 1116
DB 1332 NSAGYVDADGNKVIYDSTDNKYQAKNDGTVDTKKEVAKDK---LVAQAQTPDGTGLAQNK 1388
QY 1117 YFSLNGLQLRDAILK---NEDGTVAYYNGDGRRYENGYQFMSGVWR-----HFNNGEM 1168
DB 1389 VKSVINKEQVNDANKQGINEDNA-----FVKGLEKAAASDNKTNAAV 1431
QY 1169 SVG---LTVIDGOVYFDEMGYQAK--GKFTVTTADGKIRYFDKQSGNMYNRRIENEEGKW 1224
DB 1432 TVGOLNVAQPTPLTFAGDTGTTAKKLGSETLI-----KGGQTDITNKLTDNNIG-- 1479
QY 1225 LYLGEDGAATVGSQTINGQHLVFRANGVQVKGFEVTDHHRISYYDGHSGDQIRNRFVRN 1284
DB 1480 VVAGTDGTVTKLAKDLT-----NLNSVNAAGTKIDDK--GVSFVD--SSGQAKANTPVLS 1530
QY 1285 AQGQWYFDDNGYAVT-----GARTINGQLLYFRANGVQVKGFEVTDYGRISYYDNGSGD 1340
DB 1531 ANG---LDLGGKVISNVGKGTQD-----AANVQQLNE-VRNLLGL-----GNAGN 1573
QY 1341 QIRNRFVNAQGWYF-----DNNGYA---VTGARTI-----NGOHLVFRANGVQVK 1385
DB 1574 D-----NADGNQVNIADIKDPNSGSSNRRTVIKAGTVLGGKGNNDTEKLATGQIQVG 1626
QY 1386 GEFVTDHGRISYYDNGSGDQIRNRFVR-----NAGQWYFDDNGYAVTG 1431
DB 1627 ---VDR-----DGNANGDLSNVWKTQDKSKKALLATVNAAGQTYLTNN--PABA 1673
QY 1432 ARTINGOHL-YFRAN 1445
DB 1674 IDRINEQGIREFHN 1688
RESULT 15
US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-92LMIS:jb
; CURRENT APPLICATION NUMBER: US/11/013.759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361.619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-13

Query Match 2.8%; Score 219; DB 7; Length 1992;
Best Local Similarity 20.0%; Pred. No. 0.00013;
Matches 351; Conservative 215; Mismatches 665; Indels 524; Gaps 93;

QY 30 TISGLVKADS--NESQISNDSTSVVTANESNVI-----TEATSK 71
Db 119 TLINDLINGHAVLKEIRSSKONDKYRRITTAGSHASTAVGAMSYAOGHFSNAFGTRATAK 178
QY 72 QE-----AASQTNHVTITSSST-----SVNPKV 98
Db 179 SAYSLAVGLAATAEGOSTTAIGSDATSSSIALGALGATRAQLOGLAGSIVVTQSDN 238
QY 99 VSNPYTVGETPANGKELQNTTVDKISEAANNISKQ--TTBADTDVDDSDNAALQIL 156
Db 239 NSRPAYTPNTQALDPKFOATNNTKAGPLSIGSNSIKRKIINVAGVKNKTDVAVNAQLEAV 298
QY 157 EKLPNVKEIDGKYVYDNG--KVTNFTLADGKILHDEFAGTDSIDTVNK-DIV 212
Db 299 VWAKERRIT---FOGDDNSTDKIGLNTLTKGGA---ETNALTDNNIGVYKADNS 351
QY 213 TTRSNIYKKYNOVYDRSAOSFEHVDHYLTAESWYRPKYILKDGKTWTQSTEXDFFLLMT 272
Db 352 GLVKLAKLNNLTVNTTLNATTVTVKGSSSTTABLLSDSLTFTQNT-----402
QY 273 WMPDQETORQY-----VNYMN-----AOLGINKTYD-----TSNOL 304
Db 403 --GSQSTSVTVGVNGVKFTNNAETAAIGTTRITRDKIGFARDGDVDEKQAPYLDKKQL 460
QY 305 QLNIAAATQAKIEAKITLKNYDWLROTISAPVKTOSAWNS--DSEKPFDDHLQNGAV 361
Db 461 KVGVAITDNGIDAG-----NKKISNLAGSSANDAVTIBQLKAAKPTLNAGAG 510
QY 362 LYDNECKLTPYANSYRIILNRT--PTNQTKGKOPRYTADNTIGGYEFLLANDVDSNPVQ 420
Db 511 ISVTPTEISVDAKSG--NVTAPYNIQVKTTELNSDGTSDKFS---VKGSGTNNSLVT 563
QY 421 AEQL-NWLH-----FLMNFENIYANDPDANFDSIRVDVNDVA-DLLOIAG-----465
Db 564 AEHLASYLNEVRTADSLQSPVKEEDDDANAITVAKDTTKNAGAVSILKXKNGKLT 623
QY 466 -----DYLKAAGIKHND-----KAANDHLSILEAWSNDTPLYHDDGN 505
Db 624 VATKKGDTVTGLSQDSGLTIGKSTLNNDLGLTVKOTNEQIQV-----GANGIKFTNVNGSN 679
QY 506 -----MINMDNKLRLSLFS-----LAKPLNQRSCM---NPLITNSLVNRTDDNAETAA 551
Db 680 PTGIANTARITRDKIGFAGSGAVDTNPKPYDDQKLQVGNVKIINTGIN-AGGKAITGL 738
QY 552 VPSYFIRAHDS---EVODLIADIK---AEINPNVVGYSFTWEIEIKKAFE-IYNKDL- 603
Db 739 SPTLPSIADQSSRIELGNTIQDKOKSNAASIN-DILNTGFNLKNNNPIDFVSYDIDV 797
QY 604 -----ATEKXYTHYNALS YALLTNKSSVPRVYGDMPFDDGQYNAHKTIYEAETLL 658
Db 798 FANGNATTATVTH-DTA-----NKTS--KVWY-DVNVD-----TTIHLTGDDNK 839
QY 659 KARIKV-----SGQAMRNOQVNSE-----IITSVRYGKGALKATD 696
Db 840 KLGVTITKLKNTSANGNTATNFVNSSDEDALVNAKOIAENLTLAKEIHTTKG-----893
QY 697 TGDRTTRTSGAVIEGNPNRLKASDRVVVNMGAHQAAYRPLLLTTDNGIKAYHSDQ 756
Db 894 TADTALQITTVVKVDENN-----NADDANAITVGQKANNQVNTLTKEGELNI-----943
QY 757 EAAGLVRYNDRGELIF---TAADIKGYANPOVSGYLVGVVVPVGAALIKMFALRLARPHQ 813
Db 944 -----KTDKNGTFTFGINTTSGLKAGKSTLNDGGLSIKNPTGSEIQVGA-----DGV 991
QY 814 QNASVHON---AALD--SRVWFE--GFSNFOAPATKKEEY-----TNV---849
Db 992 KPAKYNNNGVWAGIDGTTRITRDEIGFTGNGSLDKSKPHLSKDGINAGGKKITNIQSG 1051
QY 850 VIAXN-----VDKFAEWGVTDFEMAPOY-----VSSTD 877
Db 1052 EIAQNSHDAVTCGKIYDLKTELENKISSTAKTAQNSLHFSVADEQGNFTVSNPYSSVD 1111
QY 878 GSFLDSVI-----QNGYAFTRDYDLGLISKPNKYGTADDLVKAIKA-----LHSGIKV- 925

Db 1112 TSKTSVITFAGENG--ITTKVNGKGVVR-----VGIDOTKGLTTPKLTVGNNGKGIIVD 1164
QY 926 -----MADWVPDQMYAFPEKEVVVATRVDKYGTTPVAGSQIKNTLYVVVDGKS 971
Db 1165 SONGQNTITGLSNTLANVTNDK-----GSVRTTEQGNIIKDDEKTRAASIVDVLS 1214
QY 972 SGKDOQ-----AKYGAFLELOAKYPPELFARKQISTGVPMPPS---VKIK 1014
Db 1215 AGFNLOQNGEAVDFVSTYDTVNFADGNATTAKVYDDTSTKTSKVVDVNVVDDTTIEVKDK 1274
QY 1015 QHSAXYFNGTINILGRGAG---YVLKDOAT-NYFNILSD-NKEINFL-----PKTLNQD 1063
Db 1275 KLGVK---TTTLTSTGTGANKFALSNQATGDALVKASDIAVAHLNTLSGDIQTAKASOAN 1331
QY 1064 SOVGF-SYDGKGVVYVYSTSG--YQAKNTFISGDKWYYPDNNGYMTVGAQSIG--VN 1116
Db 1332 NSAGYVDADGNKVIYDSTDNKYQAQNDGTVDKTEVAKDK---LVAQAQTPDGTTLAQMN 1388
QY 1117 YYFLSNGLQLRDAILK---NEDGTYAYYNGDGRRYENGYQFMSGVYR-----HFNNGEM 1168
Db 1389 VKSVINKEQVNDANKKQGINEDNA-----FVKGLEKAASDNKTKNAAV 1431
QY 1169 SVG--LTVIDGQVQVDEMGYQAK--GKEVTTADGKIRYFDKQSGSNMYNRNFIENEGKW 1224
Db 1432 TVGDLNVAQAQTLPTFAGDTGTAKKLGETLTI-----KGGQTDNKLTDNNIG-- 1479
QY 1225 LVLGSDGAAVTGSQTINGQHLRYFRANGVQVKGFEVTDHHRISYYDNGSGDQIRNRFVN 1284
Db 1480 VVAGTDGFTVKLAKOLT-----NLSNVNAGTKIDDK--GVSFVD--SSGQAKANTPVLS 1530
QY 1285 AQGOWFYFDNNGYAVT---GARTINGQLLYFRANGVQVKGFEVTDRYGRISYYDNGSGD 1340
Db 1531 ANG---LDLGGKVISNVGKGTKDTD-----AANVQOLNE-VRNLLGL-----GNAGN 1573
QY 1341 QTRNRFVRNAQOWFYF-----DNNGYA---VTGARTI-----NGQHLRYFRANGVQV 1385
Db 1574 D-----NADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLATGQV 1626
QY 1386 GEFVTDHGRISYYDNGSGDQIRNRFVR-----NAQOWFYFDNNGYAVTG 1431
Db 1627 ---VDK-----DGNANGDLSNVVVKTKQDKGSKALLATYNAAGQTNLTNN---PABA 1673
QY 1432 ARTINGOHL-YFRAN 1445
Db 1674 IDRINEQGIIRFHVN 1688

Search completed: February 11, 2006, 20:57:47
Job time : 19.3936 secs

THIS PAGE BLANK (USPTO)


```
Db      241  TAESVYRPKY ILKDGKWTQSTEKDFRPLLLMTWMPDQETQRYQVYVMNAQLGINKTYDDT 300
Qy      301  SNQLQNTAAATIAQIEAKITTLKNTDMLRQTIISAFVKTQAMNSDSEKPPDDHLQNGA 360
Db      301  SNQLQNTAAATIAQIEAKITTLKNTDMLRQTIISAFVKTQAMNSDSEKPPDDHLQNGA 360
Qy      361  VLYDNEGKLTYPVANSNYRLNRTPTNQTCKDPRYTADNTTGGYEFLLANDVDNSNPVVO 420
Db      361  VLYDNEGKLTYPVANSNYRLNRTPTNQTCKDPRYTADNTTGGYEFLLANDVDNSNPVVO 420
Qy      421  AEQLNLWLFNMFGNIYANDPDANFDSIRVDVNDVADLLQIAGDYLKAAGIHKNDKA 480
Db      421  AEQLNLWLFNMFGNIYANDPDANFDSIRVDVNDVADLLQIAGDYLKAAGIHKNDKA 480
Qy      481  ANDHLSILEAWSNDTPYLLHDDGNNMNDNKLRLSLFLSLAKPLNQRSGMNPPLTNSLV 540
Db      481  ANDHLSILEAWSNDTPYLLHDDGNNMNDNKLRLSLFLSLAKPLNQRSGMNPPLTNSLV 540
Qy      541  NRTDNDNAETAAPVPSYSPFIRAHDSVQDLDIADIIKAEINPNVVGYSFTWEEIKKAFIYNK 600
Db      541  NRTDNDNAETAAPVPSYSPFIRAHDSVQDLDIADIIKAEINPNVVGYSFTWEEIKKAFIYNK 600
Qy      601  DLLATEKKYTHYNTALSALLTNKSSPRVYVYGMFTDDGOYMAHKTINYEAIETLLKA 660
Db      601  DLLATEKKYTHYNTALSALLTNKSSPRVYVYGMFTDDGOYMAHKTINYEAIETLLKA 660
Qy      661  RIKYVSGGQAMNQVGNSEIITSVRYGKALKATDGTGDRTRTSGVAVIEGNNPSLRLK 720
Db      661  RIKYVSGGQAMNQVGNSEIITSVRYGKALKATDGTGDRTRTSGVAVIEGNNPSLRLK 720
Qy      721  ASDRVVNVNGAAHKNQVRPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Db      721  ASDRVVNVNGAAHKNQVRPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Qy      781  YANPQVSGYLVGVVPGVGAALIKWFALRLARPHQOMASVHQNAALDSRVMFEGFSNFQAPA 840
Db      781  YANPQVSGYLVGVVPGVGAALIKWFALRLARPHQOMASVHQNAALDSRVMFEGFSNFQAPA 840
Qy      841  TKKEBYTNVIAKNVDKFAEAGVTDPEMAPQVVSSTGDSFLDSVITQNGVAFTRDRLDLGIS 900
Db      841  TKKEBYTNVIAKNVDKFAEAGVTDPEMAPQVVSSTGDSFLDSVITQNGVAFTRDRLDLGIS 900
Qy      901  KPNKGTGADDLVKAIKALHSKGIKVMADWVPDMYAFPEKEVVTATRVKDTGTPVAGSQI 960
Db      901  KPNKGTGADDLVKAIKALHSKGIKVMADWVPDMYAFPEKEVVTATRVKDTGTPVAGSQI 960
Qy      961  KNTLVVDGKSGKQQAQYKGAFLEELQAKYPELFARKQISTGVPMDPSVKIKQWSAKY 1020
Db      961  KNTLVVDGKSGKQQAQYKGAFLEELQAKYPELFARKQISTGVPMDPSVKIKQWSAKY 1020
Qy      1021  FNGTNILGRGAGYVLKDAQNTNYFNISDNKEINFLPKTLNQDSQVGFSDGKGVVYVYST 1080
Db      1021  FNGTNILGRGAGYVLKDAQNTNYFNISDNKEINFLPKTLNQDSQVGFSDGKGVVYVYST 1080
Qy      1081  SGYQAKNTFISGDKWYFDNNGYVMTGAOSINGVNYFLSNGQLRLDAILLKNEGDGTAY 1140
Db      1081  SGYQAKNTFISGDKWYFDNNGYVMTGAOSINGVNYFLSNGQLRLDAILLKNEGDGTAY 1140
Qy      1141  YGNDGRRYENGYYQFMSGWRHFNNGEMSVGLTVIDGQVQVFDENGYQAKGKGFVTTADGK 1200
Db      1141  YGNDGRRYENGYYQFMSGWRHFNNGEMSVGLTVIDGQVQVFDENGYQAKGKGFVTTADGK 1200
Qy      1201  IRYFDKQSGNMYRNFRIENEBGKWLVLGEDGAAVTGSQTINGQHLYFRANGVQVKGEPVT 1260
Db      1201  IRYFDKQSGNMYRNFRIENEBGKWLVLGEDGAAVTGSQTINGQHLYFRANGVQVKGEPVT 1260
Qy      1261  DHGGRISYDGNSGDOIENRFRVNAQOGWFYDNNNGYAVTGARTINGQLLYFRANGVOVK 1320
Db      1261  DHGGRISYDGNSGDOIENRFRVNAQOGWFYDNNNGYAVTGARTINGQLLYFRANGVOVK 1320
Qy      1321  GEFVTDYGRISYYDGNSGDOIENRFRVNAQOGWFYDNNNGYAVTGARTINGQHLYFRAN 1380
Db      1321  GEFVTDYGRISYYDGNSGDOIENRFRVNAQOGWFYDNNNGYAVTGARTINGQHLYFRAN 1380
```

```
Qy      1381  GVQVKGFEVTDHGRISYYDGNSGDOIENRFRVNAQOGWFYDNNNGYAVTGARTINGQHL 1440
Db      1381  GVQVKGFEVTDHGRISYYDGNSGDOIENRFRVNAQOGWFYDNNNGYAVTGARTINGQHL 1440
Qy      1441  YFRANGVOVKGEFVTDYGRISYYDANSGERVRIN 1475
Db      1441  YFRANGVOVKGEFVTDYGRISYYDANSGERVRIN 1475

RESULT 2
JT0345
debranchase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C:Accession: JT0345; C31335
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: JT0345; MUID:89137980; PMID:2976010
A:Accession: JT0345
A:Molecule type: DNA
A:Residues: 1-1375 <UED>
A:Experimental source: UNIPROT:P13470; UNIPARC:UPI0000155515
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A3135; MUID:87308013; PMID:3040685
A:Accession: C31335
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: UNIPARC:UPI000014E25D; GB:M17361
A:Gene: gtfC
C:Function:
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
A:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-1375/Product: glucosyltransferase #status predicted <MAT>
F:1126-1145/Domain: cpl repeat homology <CP1>
F:1123-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match      68.3%; Score 5285; DB 2; Length 1375;
Best Local Similarity 75.9%; Pred. No. 3.4e-248;
Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;

Qy      1  MDKKVRYKLRKVKRWVTVSVASAVMTLTTLSGGLVKADSNESKSIQINDSNTSVVTANE 60
Db      1  MEKKVRFKLRKVKRWVTVSIASAVVTLTSLSGSLVKADSTDDRQQAFTESQASLVTTSE 60
Qy      61  -----ESNVLTATSQEAASSQNTHTVTTSSSTSVVNPKEVSNPVTVCETA-- 109
Db      61  AAKETLTATDTSTATSFATSA--TVTDVSTTNSQNTTANTANTANFVVKPTTSEQAKT 118
Qy      110  SNGEKLQNTTVDKTSSEA-----AANNISKQTTEAD-----TDVIDDSN 149
Db      119  DNSDKIITTSKAVNRLTATGKFPVANNNTAHPKVTDKIVPIKPKIKLQPSLSQDDI 178
Qy      150  AANIQLIEKLPNVKEIDGKYYYNNNGKVRNFTLIADGKILHFDGTGAYTDTSDTVNK 209
Db      179  AA----LGNVKNIRKVKNGKYVYKEDGTLQKNYALNINNGKTFPFFDETGALESNTLPSKKG 234
Qy      210  DIVTT-RSNLYKKYNOVDRSAQSFHVHVLTAESWYRPKYLKDGKTKTQSTKDRPP 268
Db      235  NITNDNTNFAQNVYVSTVDVANFHVHDLTAESWYRPKYLKDGKTKTQSTKDRFP 294
Qy      269  LLMTWVDPQETQRYQVYVMNAQLGINKTYDDTNSQLQNLNIAAATIAQIEAKITTLKNTD 328
Db      295  LLMTWVDPQETQRYQVYVMNAQLGINKTYDDTNSQLNLAATQITQTKIEEKITAENKTN 354
```

329 WLRQTISAFVKQTQSANWSSEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRIILNRTPTNQT 388
355 WLRQTISAFVKQTQSANWSSEKPFDDHLQKGLLYSNNKSLTQSANSYRIILNRTPTNQT 414
389 GKQDPRTADNTTGGYEFLLANDVNSNPVQAEQLNLWFLMNFNGNIYANDPDANFDSI 448
415 GKQDPRTADRTTGGYEFLLANDVNSNPVQAEQLNLWFLMNFNGNIYANDPDANFDSI 474
449 RVDADVNDADLLQIAGDYLLKAAKGIHKNDDKAANDHLSILEAWSNDTPLYLHDDGDNMTN 508
475 RVDADVNDADLLQIAGDYLLKAAKGIHKNDDKAANDHLSILEAWSNDTPLYLHDDGDNMTN 534
509 MNKURLSLFLSLAKPLNQRSGMNPILITNSLVNRTDDNAETAAPVPSYSFIRAHDSVQDL 568
535 MDNRLRLSLLYSLAKPLNQRSGMNPILITNSLVNRTDDNAETAAPVPSYSFIRAHDSVQDL 594
569 IADIILKAEINPNVVGYSFTMEBIKKAFETYNKDLLATEKKYTHYNTALSYALLTNKSSV 628
595 IIRNIIRTEINPNVVGYSFTTEBIKKAFETYNKDLLATEKKYTHYNTALSYALLTNKSSV 654
629 PRVYVGMFTDQGYMAHKTINYEALTEILLKARIKYVSGQAMRNOQVGNSEIITSVRYG 688
655 PRVYVGMFTDQGYMAHKTINYEALTEILLKARIKYVSGQAMRNOQVGNSEIITSVRYG 714
689 KGALKATDGTDRTRTSGVAVIEGNNPSSLKASDRVVVNMGAHKNQAYRPLLLTTDNG 748
715 KGALKATDGTDRTRTSGVAVIEGNNPSSLKASDRVVVNMGAHKNQAYRPLLLTTDNG 774
749 IKAYHSDQBAAGLVRYTNRGELIFTAADI KGYANPQVSGYLGWVVPVGAALIKMPALRL 808
775 IKAYHSDQBAAGLVRYTNRGELIFTAADI KGYANPQVSGYLGWVVPVGAADQ - DVRV 832
809 A - - - RPHQMAVSHVMAALDSRVMEGFSNFQAFATKKEEYTNVVIKKNVDKFAEWGVD 865
833 AASTAPSTDCKSVHQNAALDSRVMEGFSNFQAFATKKEEYTNVVIKKNVDKFAEWGVD 892
866 FEMAPQVSVSTDSGLDSVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKV 925
893 FEMAPQVSVSTDSGLDSVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKV 952
926 MADWVPDQMYALPEKEVVTATRVDKYTPVAGSQIKNTLYVVDGKSSGKQQAQYGGAPL 985
953 MADWVPDQMYALPEKEVVTATRVDKYTPVAGSQIKNTLYVVDGKSSGKQQAQYGGAPL 1012
986 BELQAKYPELFARKOISTGVPMDPSVKIKQWSAKYFNGNIIILGRGAYLVKDOATNTYEN 1045
1013 BELQAKYPELFARKOISTGVPMDPSVKIKQWSAKYFNGNIIILGRGAYLVKDOATNTYPS 1072
1046 -ISDNKEINFLPKTLN-----QDSQVGFSDGKGYVYSTSGYQAKNTFISEGDKWYVF 1099
1073 LVSDN---TFLLPKSLVNPNGHGTSSSVTGLVFDGKGYVYSTSGNQAQNAFISLGNWYVF 1129
1100 DNNGYMTGAQINGVNYVFLSNGIQLRDAILKNEBDGTVAAYGNDGRRYENGYYOYFMGCV 1159
1130 DNNGYMTGAQINGVNYVFLSNGIQLRNAIYDNGKNVLSYYGNDGRRYENGYYILF-GQQ 1188
1160 WRHFNGNEMSVGLTWIDGOVQYFDEMGYQAKKFVTTADGKIRYFDKQSGNMRYRFTEN 1219
1189 WRYFQNGIMAVGLTRVHGAVQYFDEMGYQAKKFVTTADGKIRYFDKQSGNMRYRFTEN 1248
1220 BEGKVLILGEDAAVTGSGTINGOHLVFRANGVQVKGFEVTPDHGRISYDNGSGDOIRN 1279
1249 SKGEWFLPDHNGVAVTGTFTNGQRLYFKPNGVQAQGEFIRANGVLYRYDNPNSNEVRN 1308
1280 RFRVNAQQGWYFDDNNGVAVTGARTING-----QLLYFR 1313
1309 RFRVNSKGEWFLPDHNGVAVTGTFTNGQRLYFKPNGVQAQGEFIRANGVLYRYDNPNSNEVRN 1347

RESULT 3
A38175
glucosyltransferase precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus
C>Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004
C:Accession: A38175
R:Abn. H.: Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A:Reference number: A38175; MUID:91123227; PMID:1704006
A:Accession: A38175
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1592 <ABO>
A:Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:G217032; PIDN:BAAL4241.1; PID
F:1093-1112/Domain: cpl repeat homology <CP1>
F:1222-1241/Domain: cpl repeat homology <CP2>
F:1287-1306/Domain: cpl repeat homology <CP3>
F:1330-1351/Domain: cpl repeat homology <CP4>
F:1352-1371/Domain: cpl repeat homology <CP5>
F:1402-1420/Domain: cpl repeat homology <CP6>
F:1465-1484/Domain: cpl repeat homology <CP7>
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 58.5%; Score 4531; DB 2; Length 1592;
Best Local Similarity 55.8%; Pred. No. 1.5e-211;
Matches 889; Conservative 213; Mismatches 344; Indels 146; Gaps 19;
QY 1 MDKVRKYLKRVKRWTVSVASAVMTLTTLTSGGLVKADSNESKSOISNDSNTSVVTANE 60
DB 1 MEKNVRFKMKVKRWTVLSVASATMLASALGASASADTDTA-----SDSNQTVVTGQ 56
QY 61 ESNVTEATSKQEAASSQTNHTVTSSTSSSTSVVNPKE-----VVSNPYTVGETASNGEKL- 115
DB 57 TTN--NQATDQTSIAATATSEQASTDAATDQASAAEQGTGTTASTDTAAQTITNANEAK 114
QY 116 -----NQQTTVTKTSEAAANNISKQTEADTVDIDSNANLQILEKLPVKEIDGKY 169
DB 115 WYPTENENQGFDEMLAEA-----KNVATSEDSIPSD-----LAKSNVKQVDGKY 161

QY 170 YYYDNNGKVRNTFTLITADGKILHFDGTGAYTDTSDTIVNK--DIVTTRNLNKKYKNQVYD 227
DB 162 YYYDQDGNVKNFVSVGDKIYVFDGTGAYKDTSKVDADKSSVASQNAITFAANNRAVS 221
QY 228 RSAQFEHVDHYLTAEVSRPKYILKDGKTWTQSTEKQPRLLMTWPPQDETQRYVNTM 287
DB 222 TSAKNFEADVNYLTADSVWRPKSILKDGKTWTEGKDDFRPLMAWVPDTETKRYVNTM 281
QY 288 NAQLGINTYDDTSNQLQNTAAATIOAKIEAKITTLKNTDWLRQTIISAFVKTQSNWSD 347
DB 282 NKVVGIDKTYTAETSQADLTAAAEVLQARIQKQITSENNTKWLREAIISAFVKTQPMNGE 341
QY 348 SEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRIILNRTPTNQTGCKDPRYT--ADNTIGGYE 405
DB 342 SEKPFDDHLQNGALLFDNQDTLTPDTQSNYRLNRTPTNQTGSLDSRFTYNNPNDPLGGVD 401
QY 406 FLLANDVNSNPVQAEQLNLWFLMNFNGNIYANDPDANFDSIRVDADVNDADLLQIAG 465
DB 402 FLLANDVNSNPVQAEQLNLWFLMNFNGSIYANDADANFDSIRVDAEDVNDADQLQISS 461
QY 466 DYLKAAKGIHKNDDKAANDHLSILEAWSNDTPLYLHDDGDNMTNMDNKLRLSLFLSLAKPL 525
DB 462 DYLKAAAGIDKKNKNVANNHVSIVEAWSNDTPLYLHDDGDNLMNMDNKFSLMWSLAKPT 521
QY 526 NORSQMPNLITNSLVNRTDDNAETAAPVPSYSFIRAHDSVQDLIADIKAETINPNVVGYS 585
DB 522 DVSRLNPLIHLNLSLVNRTDREVETVPSYSPARAHDSVQDLIADIKAETINPNVVGYS 581
QY 586 FTMBEIKKAFETYNKDLLATEKKYTHYNTALSYALLTNKSSVPRVYVGMFTDQGYMA 645
DB 582 FTQEEIDQAFKIYNEIDLKSKDKKYTHYVPLSYTLTLLTNKSGIPRVYVGMFTDQGYMA 641
QY 646 HKTINYEAETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALKATDGTDRTRTS 705
DB 642 NKTVNYDAIESLLKRMKYVAGQAMQNYQIGNGEILTTSVRYGKALKQSDKGDAATTRTS 701

```
QY 706 GVAVIEGNPSLRKASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
Db 702 GGVVMGNQNPSPSLDGK-VVALNMGAHANQBYRALMVSTKDGVAITYATDADASKAGLVK 760
QY 764 YTNDRGELIFTAADIKGYPANPOVSGYLVGVVPGAA---LIMFALRLARPHQWASVHQ 820
Db 761 RTDENGILYFLNDDLKGVPANPOVSGFLQVVPVGAADDQDQIRVAASDTASTDGK--SLHQ 818
QY 821 NAALDSRVMEFSGFNFQAPATKEBYTNVIANQVDKPAEAGVDTFEMAPQVVSSTDSGF 880
Db 819 DAAMDSRVMEFSGFNFQAPATKEBYTNVIANQVDKPAEAGVDTFEMAPQVVSSTDSGF 878
QY 881 LDSVLQNGVAFTRDYLGISKNKYGTADLLKALHSHKGIKYNADVPDQWAFPEK 940
Db 879 QBSVIONGVAFTRDYLGLSKANKYGTADQLKALHSHKGLKYNADVPDQWAFPEK 938
QY 941 EVVTRATRVKYGTPVAGSGQIKNTLVVDGSKGQQAQYGGAFLEELQAKYPELFAFKQ 1000
Db 939 EVVTRTRDKFGKPIAGSQINHSLVYDTKSSGDDYQAKYGGAFLEELQAKYPELFAFKQ 998
QY 1001 ISTGVPMPSVKIKQWAKYFNGTILGRGAGYVLKQDQATNTYFNI SDNKEINFLPKTLL 1060
Db 999 MSTQAIDPSPVKIKQWAKYFNGTILGRGAGYVLKQDQATNTYFNI SDNKEINFLPKTLL 1056
QY 1061 NQDSQVGFSDGKGVVYST-SGYQAKNTPFISEGDKWYFEDNNGYMTGAQSGINGVNYF 1119
Db 1057 GKWESGIRYDQKGVYINSSATGDQVQKAFITEAGNLVYFGDKGVYMTGAQSGINGVNYF 1116
QY 1120 LNSGLQRLDALIKNEDGTVAAYGNDGRY--ENGYQFMGSGWRHFNNGEMSVGLTVTDG 1177
Db 1117 LENGTALENTIYTDAGNSHYANDGKYENENGQQP-GNDWRVFKDGNMAVGLTVDG 1175
QY 1178 QVQYFDEMGYQAKGFVTTADKIRYFDKQSGNMYRNRFIENEGKWLVLGDEGAAVTGS 1237
Db 1176 NVQYFDDKGVQAKDKIIITRDGKRVYFDQHNGNAVNTNFIADKTGHVYLGKGVAVTGA 1235
QY 1238 QTINGQHLVFRANGVQVKGFEVTDHGRISYYDGHSGDQIRNRFVRNAQGWYFDDNGY 1297
Db 1236 QTVGQKLYFEANGQVQKGFVTSDEGKLYFYDVSDDGMDWTDFTIEDKAGNMYFLGKQGA 1295
QY 1298 AVTGARTINGQLLYFRANGVQVKGFEVTDHGRISYYDGHSGDQIRNRFVRNAQGWYF 1357
Db 1296 AVTGARTINGQLLYFRANGVQVKGFEVTDHGRISYYDGHSGDQIRNRFVRNAQGWYF 1355
QY 1358 DNGYAV-----TGAR 1368
Db 1356 GNDGVADPSPVVKGTQFKDASGALRYNLKGQLVTGSGWYETANHDWVYIQSGKALTGEQ 1415
QY 1369 TINGQHLVFRANGVQVKGFEVTDHGRISYYDGHSGDQIRNRFV----- 1412
Db 1416 TINGQHLVFKDGHQVKGQLVTGTDGKRVYVDANSQDQAFNKSVTNGKTYFYGNDGTAQ 1475
QY 1413 -----RNAQGWYFDDNGYAVTGARTINGQHL 1440
Db 1476 TAGNPKGTQFKDGSIRFYSMEGQLVTGSGWYVNAQGWLYV-KNGKVLTLGLQTVGSQRV 1534
QY 1441 YFRANGVQVKGFEVTDHGRISYYDANSGERV 1472
Db 1535 YFDENGIQAKGKAVRTSDGKIRYFDENSGSMI 1566

RESULT 4
A:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A:Reference number: A45866; MUID:91100958; PMID:2148600
A:Accession: A45866
A>Status: preliminary
```

```
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 49.6%; Score 3839.5; DB 2; Length 1431;
Best Local Similarity 53.1%; Pred. No. 4.2e-178;
Matches 765; Conservative 231; Mismatches 393; Indels 53; Gaps 22;

QY 1 MDKRVYKLRVKKRWTVSVASAVMTL-TTTLGGGLVKADSNESK-----QISNDSNTS 54
Db 1 METKRYKWKHKWVTVAVASGLITLGTTLGSSVSAETEQDTSKVVTKQSEDDKAA 60
QY 55 VVTANEESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKEVSVNPYTVGETASNGEK 114
Db 61 SESSQTDAPKTKQAQTEQTQAQSQAN-VADTSTSTIKETPSQNIITQAQNSDDKTVTNTKS 119
QY 115 LQNTQTV-DKTSEAAANNISKOT-TEADTVDIDSNAA-----NLQILEKLPNKVEIDG 167
Db 120 EEAQTSERTKQSEEAQTASSQALTKQAKBLTKQRTAAQENKPNVDLAAIPNKKQIDG 179
QY 168 KYVYDNGKVRNFTLIADGKIHLHFD-TGAYTDTSDTVNKOIVTTRSNLYKKYNQVY 226
Db 180 KYVYISGQPKKFPALTVNNKLVYDKNTGALTDTTS-QYQFKQGLTKLNNDYTHNQIV 238
QY 227 DRSAQSPHVDHYLTAEWSYRPKVILKDGKWTOSTEKDFRPLMTWMPDQBTQRYVNY 286
Db 239 NFENTSLETINYVTADSWYRPKDILKNGKTWTASSESDLRPLLSWMPDKQTQIAYLNY 298
QY 287 MNAQ-LGINKTYDDTSNQLNIAAATQAKIEAKITTLKNTDMLRQITSAFVKTSQAWN 345
Db 299 MNQOGLGTGENYTDSSQESLNLAQTQVQKIEKISQTTQQTQWLRLDIINSFVKTPQNNN 358
QY 346 SDSEKPPD---DHLQNGAVLYDNEGKLTVPYANSNYRILNRTPTNQTCKDPRYTADNTI 401
Db 359 SQTSDTSAGKEDHLOGGALLYNSDK-TAYANSYRLNLRTPTSQTK--PKYFEDNS 415
QY 402 GGYEFLANDVDSNPVQAEQLNWLHFLMFGNITANDPDANFDSIRVDAVDNVDADLL 461
Db 416 GGYDFLLANDIDNSNPVQAEQLNWLHFLMNYGSIIVANDPEANFDGVRVDAVDNVDADLL 475
QY 462 QIAGDYLKAAKGIHKNKDKAANDHLSILEAWSNDPTPLHDDGDNNMINMDNKLRLSL 521
Db 476 QIASDYLKAHYGVDPKSEKNAIHNLSILEAWSNDPQYNKDKTGAQLPINDKLRLSLVAL 535
QY 522 AKPLNQ-----RSGMNPILTNLVNRTDDNAETAAPVPSYFIRAHDSVEQDLIADI 573
Db 536 TRPLEKASKNKNERISGLEPVTISLNNRSEKNSERMANYIFIRAHDSVEQTVIAKII 595
QY 574 KAEINPNVVGYSFTMEEIKKAFIYNKOLLATEKKYTHYNTALSAYALLTNKSSVPRVY 633
Db 596 KQAINPKTDGLTFTLDELKQAFKIVNEDMRQAKKYYTQSNIPTAYALMLSNKDSITRLYY 655
QY 634 GDMFTDDQYMAHKTINTYEIETLLKARIKIVSGQAMRNQOVGNSE-----IITS 684
Db 656 GDMYSDDDQYMATKSPYYDAITLLKARIKAAAGGQDMKITTYVEGDGSHMDWDYTGVLTS 715
QY 685 VRYGKALKATDGTGRTTTSQVAVIEGNNPSRLRLKASDRVVNMGAHKNQAYRPLLT 744
Db 716 VRYGTGANEATDQGSSEATKTOGMVITSNNSLKNQNDKVIIVNMGAAHKNQAYRPLLT 775
QY 745 TDNGIKAYHSDQEAAGLVRYTRNDRGELIFTAADIKGYANPOVSGYLVGVVPGVGAALIKMF 804
Db 776 TKDGLTSYTSDAAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLVGVVPGVGA--DNQ 833
```



```
Db 1219 YFDNGYQAGVFPVPTKDGHLMPFCGDSGERKYSGFEE-QDGNWYVANDKGYVATGFTKV 1277
QY 1236 -----GSGTINQHLFRANGVQVKGFEV 1259
Db 1278 GKQNLFPNEKGVQVKNRPFQVGDATYANNEGVDLRGAQTINGDELYFDESQVKGFEV 1337
QY 1260 TDHGHISYDGNSGDQIRNR-----FVRNAQGWQFYFDNN 1295
Db 1338 NNPDGTTSYDAITGVKLVDTSLVVDGQTFNVDAKGVTTKAHTPGFYTTGDNWFFYADSV 1397
QY 1296 GYAVTCAPTINGQLLYFRANGVQVKGFEVTDYGRISYDGNSGDQIRNRFVRNAQGWQF 1355
Db 1398 GRNVTGAQVINGQHLFYDANGQVKGFEVNTDGSRSFYHWTGDKLVSTFFATGHDWY 1457
QY 1356 YFDNNGYAVTGARTINGQHLFRANGVQVKGFEVTDYGRISYDGNSGDQIRNRFVRNA 1415
Db 1458 YADDRGVVTVGAQVINGQKLFDTGKQVKGAFATNANGRSYHYHNTGKLVSTFFATG 1517
QY 1416 QGWQFYFDNNGYAVTGARTINGQHLFRANGVQVKGFEVTDYGRISYDGNSGDQIRN 1471
Db 1518 DNNWYADAKGVVVGEQINGQHLFYDQTKQVKGATATNPDGSIYYDVHTGEK 1573

RESULT 7
T31098
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31098
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.
FEMS Microbiol. Lett. 159, 307-315, 1998
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DS
A:Reference number: Z20981; MUID:98164374; PMID:9503626
A:Accession: T31098
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1508 fMON>
A:Cross-references: UNIPROT:052224; UNIPARC:UPI00000BB69B; EMBL:AF030129; NID:G2766611;
A:Experimental source: strain NRRL B-1299
C:Genetics:
A:Gene: dsrB
C:Function:
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 41.9%; Score 3243; DB 2; Length 1508;
Best Local Similarity 43.7%; Pred. No. 3.7e-149;
Matches 682; Conservative 241; Mismatches 452; Indels 184; Gaps 32;

QY 2 DKKVRKLRKVRKRWTVTS-VASAVMTLTTLGGLVKADSNESKSQ-----ISNDSTSVV 56
Db 6 ERNRKLYKSKSVWIGGLISTLMTSQTQNVNADSTNTVTDKSVTVSSNNTT-- 63
QY 57 TANEESNVITEATSKQEAASSQTNHTVTTSSTSVNPKYVSNPYTVGETASNGEK-- 114
Db 64 --NQHDTVVDKQTI--PVKNDQTTQIIAANAQAQKVKASDTTDTQKQAEATANNTKED 119
QY 115 -LQNOT-----TTVDKTSAAANN----- 132
Db 120 SIDNLTQKLPVPTANQKTGYLEKDGKWWYVTSNTLAKGLTTVDNHHKQYFDNNGVQAK 179
QY 133 -----ISKQT-----TEADTDVDDSN 149
Db 180 GQFVTDNCKTYLDPNSGNVATGIIQIGSQTALFNDNGEQVPADPYTAPDGKTYTFDDKG 239
QY 150 AANLQILEKLPNVKEIDGKYIYDNNKQVNTFTLIADGKILHFDGTAYTDTSDTVDNK 209
Db 240 QATI-----GLKAINGHYTFDSLGQLKKGFTGVIDQVRYFDQESQGVSTTDSQIK 292
QY 210 DIVTTRSNLYKKYNOVDRSAQSFHVDHLYFAESWYRPKYLLDKGKTWTOSTEKDPRPL 269
Db 293 EGLTSQTTDYTAHNAVHSDSADFONFNGYLTASSWYRPKDVLRNGQHWHEATTANDFRPI 352
```

```
QY 270 LMTWPPDQETQROYVNYMNAQLGI---NKTYDDTSNQLQNLIAAAATQAKIEAKITTLKN 326
Db 353 VSWWFPKSTQVNYLYNM-SQMGLDNRQMFSLKQNAMLNACTTVOQAIEKIGVANS 411
QY 327 TQWLRTQISAFVKTSQAMNSDEKPPDDHLQNGAVLYDNEKGLTPYANSNYRLNRPPTN 386
Db 412 TAWLATAIDDFIRTQPMNMSSEDPKNDHLQNGALTFFVN-SPLTPDTNSNFRLLNRPPTN 470
QY 387 QTKGKDPRTADNTTGGYEFILANVDNSNPVQAEOLNWLHFLNMFNFIYANDPDPAED 446
Db 471 QTVG--PKYTIQDSKGGFELLANDVDNSNPVQSQNLNWLHFLNMFNFGSITANDSAANFD 528
QY 447 STRDAVDNVADADLQIAGDYLKAAKGIHKNDKAAANDHLSILEAWGNDNTPYLLHDDGDNM 506
Db 529 GIRDAVDNVADADLQIADADYFKAAYGVNDKNDATANQHLISILEDWSHNDPEYVKDFGNQ 588
QY 507 INMDNKLRLSLFSLAKPLNQRSGMNPILTNSLVNRTDDNAETAAVPSYSFIRAHDSEVQ 566
Db 589 LTMDDYMHQTQLIWSLTCKMRMGTTQRFMDYYLVNRRHDS TENTAIPTNYSFVRAHDSEVQ 648
QY 567 DLIAIDIIKAEINPVV-GYSFTMEBIKKAFEIYNKDLATEKKYTHYNTALSVALLLTNK 625
Db 649 TVIAQII-SELHPDVKNSLAPTADQLAEAFKTYNNDKQADKKYTOYNNPMSAYAMLLTNK 707
QY 626 SSVPRVYVYGDMMFTDDGQYMAHKTINYEATETLLKARIKYVSGQAMRNQOVGSEIITSV 685
Db 708 DTVPRVYVYGDLYDDGQYMAHKSPPYFDALINGLLKSRIKYVAGQSM---AVDQNDILTNV 764
QY 686 RYKGALKATDGTDRTRTSGVAVIEGNNPNSURLKASDRVVVNMGAHQAQYRPLLLTT 745
Db 765 RYKGAMSVTSGNADTRTQIGVIVSNKENLALKSGDVTVLHMGAAHQAQYRPLLLGT 824
QY 746 DNGIKAYHSDQEAAGLVYTDNRGELIPTAADIKGYANPOVSGYLGWVWVPGVGAALIKMFA 805
Db 825 ADNLSYYDNDNAP---VKYTNDOGLIIFDNTIYGVNRPQVSGFLAVWVPGVAD-----S 876
QY 806 LRLAR-----PHQOMASVHONAAALDSRVMPGFSNFQAFATKCEYTNVWIAKNVDKFA 859
Db 877 HQDARTLSDDTAHDGKTFHSHNALLDSQVIEGFSNFQAFATNTEDYTNAVIAKNGQLPK 936
QY 860 EMGVTDPEMAPQYVSSSTDGSLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIALH 919
Db 937 DWGITSFQLAPQYRSSTDTSLDSIIQNGYAFTRDYDLGYGTPTKYGTVDQLRDAIKALH 996
QY 920 SKGIKWADWPDQMYAFPEKEVVTATRVDKYCTPVAGSQIKNTLYVVDGKSSGKQQAQ 979
Db 997 ANGIOAIADWPDQIYNLPQGLATVTRTNSYGDKTNSDIDQSLYVQSRGGK-YQAQ 1055
QY 980 YGGAFLBELQAKYPELPARKOISTGVPMDDPSVKIKQWSAKYFNGTNILGRGAGYVLKQQA 1039
Db 1056 YGGAFLSDIQKYPALFETKQISTGLPMDPSQKITEWSGKYFNGSNIQKAGYVLKDSG 1115
QY 1040 TNTYFNI-SDNKEINFLPKTLINQDSQVGSYDGGYVYVYSTGYQAKTFISEG-KWY 1097
Db 1116 TDQYKVTSSNNRRDFELPKQLTDDLSETFGVDRDNGWVYVYLSGLYLRNTPFQDNGNY 1175
QY 1098 YPDNNGYVWVTAQSGINGVNYVFLSNGLOLRDAILKNEDGTAYAYNGDGRYENGYYQFWS 1157
Db 1176 YPDSGHLVTFQNIINHHYFPLNGIELVQSLQNVADGSTTYFDQKGRQVFNQYITDOT 1235
QY 1158 GWRHFF-NNGEM-SVGLTVIDQVQYFDBMGVQAKGFVTTADTKIRYFDKSGMYNRN 1215
Db 1236 GTAYFQNDGTWTSFGFTTEIDGHKQFYKNGQVQKQFVSDTDGHVFLYEAGNGVATOR 1295
QY 1216 FIENEKWLXILGEDAAVTSQTINGQHLFRANGVQVKGFEVTDYGRISYDGNSGDQIRN 1263
Db 1296 FAQNSQGWFLYDNDGIALTGLQTINGQVNYFYADGHQSKGDFITIQNHVLTNPLTGA 1355
QY 1264 -----GRISYDGNSGDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQLLYFRAN 1315
Db 1356 TTGMOQIGDKIFVFD-NTGNMLTNOYQTLDDQWLHLSTQGPADTGLVNLINGNLKYFQAN 1414
```

QY 1316 GVOVKGEFVDRYGRISYY-DGNSGDOIIRNFRVNAQOMGFYDNNGYAVTGARTINGQH 1374
Db 1415 GRQVKGFVTDPTITVNSYMTATDSAVFNDYF-TYQGMVLTDSNYQLVKGFKVNNKL 1473
QY 1375 LYF-RANGVQVKGEFVTDHRGRISYYDGNSGDOIIRNFRVNAQOMGFYDNNGYAVTGCA 1432
Db 1474 QHFEITGVQYKSAHII-----VNNR-----TYIFDDQGYFVSVA 1508
RESULT 8
JC5473
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C:Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
C:Accession: JC5473
R:Monchois, V.; Willemot, R.M.; Rемаud-Simeon, M.; Croux, C.; Monsan, P.
Gene 182, 23-32, 1996
A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leucon
A:Reference number: JC5473
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1290 <MON>
A:Cross-references: UNIPROT:Q48756; UNIPARC:UPI000017ABDF; GB:U38181
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose on
C:Genetics:
A:Gene: dsra
C:Keywords: glycosyltransferase; hexosyltransferase
F:78-870/Domain: catalytic #status predicted <CAT>
F:922-1290/Domain: glucan-binding #status predicted <GB>
Query Match 41.9%; Score 3242; DB 2; Length 1290;
Best Local Similarity 49.7%; Pred. No. 3.2e-149;
Matches 664; Conservative 203; Mismatches 401; Indels 68; Gaps 23;
QY 159 LPNKHIDKYYDYDNGKVRNFTLIADGKILHFD-ETGAYTDSITVNKDIYTTSEN 217
Db 1 MPNVKHVDGKVFYFGDDGPKNFNTIIDGPKYFDFKDTGALSNN-----DKQYVSELS 55
QY 218 LYKYNQVYDSAOSEFHVHYLTAEWSYRKYILKDGKWTQTSTQEKDFRPLMTWPDQ 277
Db 56 IGNKHNAVYNTSSDNFTQLEGHLTASSWYRPKDILKNGKRWAPSTVTDFRPLMAWWDK 115
QY 278 ETQRYVYMAQLGINT--YDTSNQLQINIAAATQAKIEKITTLKNFDMLRQRTIS 335
Db 116 STQVTLNMYKMQGLSGTHFSDNENRMTLTAAAMQAQVNIKKIGQLGNTDMLKTAMT 175
QY 336 APVKTQSAWNSDSEKPFDDHLQNGAVLYDNEGKLTTPYANSYRILNRTPTNOTGKKDPY 395
Db 176 QYIDAQPNWIDSEAKGDDHLQGGALLYN-SDMSPKANSYRKLSRTPKXKQGIADKY 234
QY 396 TADNTIGYEFLANDVNSPVPVQAEQLNMLHFLMNFNGIYANDPDANFDSIRVDAYDN 455
Db 235 KQ----GGFELLANDVNSPVPVQAEQLNMLHYMWNIGSILQNDQANFDGYRVDAYDN 290
QY 456 VDADLLOIAGDYLLKAAGIHNKDKAANDHLSILEAWSND-----TPYLHDDGDNMINMDN 511
Db 291 VDADLLOIAGDYLLKAAGIHNKDKAANDHLSILEAWSND-----TPYLHDDGDNMINMDN 346
QY 512 KLRSLFSLAKPLNQRSGMPLNTNSLVNRTDNDNAETAAPVSPYSFIRAHSEVDLTIAD 571
Db 347 PLHLAIKATVNLNPNDRSGLEPTEHSLVKRITDDEKENVQPNYSFIRAHSEVDTIAD 406
QY 572 IIKAEINPNVVG--SFTMEIKKAFIYNKDLLATEKKYTHYNTALSYALLLTKSSVP 629
Db 407 IIKKINPASTGLDSTVTLQIKQAFDIYNADLAKADKYTPYINIPASVALLLPNKDTP 466
QY 630 RYVYGDMTDDQYMAHTIYEAETILLKARIKTVSGQAMR-NQQVGNSEIITSVRYG 688
Db 467 RYVYGDMTDDQYMAKQSPYYQAIADALLKARIKTVSGQAMR-NQQVGNSEIITSVRYG 526
QY 689 KGALKATDTGRTTSTGVAIVIEGNNPSLRKASDRVVVNMGAHKNQAYRPLLLTTONG 748

Db 527 KGAMTASDSGNQETRYQGI GLVNNR PDLKLSDKDEVMGMAAHKNQDYPVLITTKSG 586
QY 749 IKAYHSDOEAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVWPVGAALIKNFALR 808
Db 587 LKVYSTDANAP-VVR-TDANGOLTFKADVMYGVNDPQVSGYIAAWVPVGAASENQDARTKS 644
QY 809 ARPHQOMASV-HQNAALDSRVMPFSGFNFQAFATKKEEYTNVVIKAKNDKFAEAGVTDPE 867
Db 645 ETTQSTGVSVHSNAALDSQVIEGFSNFDPPTTDEFITNIKIAQNVLNPKDWGITSPE 704
QY 868 MAPQVSVSTGSLFSDSVIQNGYAFTRDYDLIGISKPNKYGTADDLKAALKALHSKGIKMA 927
Db 705 MAPQVRASSDSFLDAIVQNGYAFTRDYDIGNTPTKGTADNLDDLALHGGQIQAIN 764
QY 928 DWVPDQMYAPPEKEVVTATRVKYGTVPVAGSQIKNTLYVVDGKSSGKQQAQYKGFAPLEE 987
Db 765 DWVPDQIYNLPDEQLVTAIRTDGSDHTYGSVIDHTLYASKTVAGGIYQO-QYCGAFLEQ 823
QY 988 LOAKYPELFARKQISTGVPMDDPSVKIKOWSAKYFNGTNILGRGAGYVLKDAQNTYFNIS 1047
Db 824 LKTYQPLFQKQISTDQPMNPDIIQIKSWEAKYFNGSNIQGRGAWYVLKDWGTQOYFNVS 883
QY 1048 DNKEINFLPKTLNMODSOVGFSDGKGYVYYSTSGYQAKNTFISEGDKWYTFDNNGYMT 1107
Db 884 DAQ--TFLPKQLLGEKAKTGFTVRGKETSFTSTGYQAKSAFICDNGWYTFDDKGMV 941
QY 1108 GAQINGVNYFLNSLQRLDAILKNEGTAYYNGDGRYENGYYQFMSGVWRH-FNNG 1166
Db 942 GNQVINGINYFLPNGLIELQDAYLVH-DCMYVYNNIKQLHNTYVYQDKQNFHYFFEDG 1000
QY 1167 EMSVGLTVI---DQO--VOYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYNRRIENEE 1221
Db 1001 HMAQGIVTIIQSDDGTPVTQYFDENGKQKQGVAVXGSDGHLHYFDGASGNLFPKSGRLAD 1060
QY 1222 GKWLYLGDGAATGSGOTINGOHLVFRANGVQVKGEFVTDHHRISYYDNGSGDOIIRNRF 1281
Db 1061 GSWLYDEKGNVATGKTIINQTVF-----NDDGRQIKNPF 1097
QY 1282 VRNAQGVFYDNNGYAVTGARTINGQLLYFRANGVQVKGEFVTDYGRISYYDNGSGDQ 1341
Db 1098 KELADGSLYLNKNGVAVTGQIIINGQTLFPGNDGRQPKGTTHINATGESRYDPSGNM 1157
QY 1342 IRNFRVNAQOMGFYDNNGYAVTGARTINGOHLVFRANGVQVKGEFVTDHGRISYYDQ 1401
Db 1158 ITDRFERVGDNQWAFYGYDGVAVTGDRITIKGNLYFNQNGIQMKGHLRL-NGIMRYDDA 1216
QY 1402 NSGDOIIRNFRVNAQOMGFYDNNGYAVTGARTINGOHLVFRANGVQVKGE---FVTDY 1458
Db 1217 DTGELVRNFRVLLSDGSHVYFGDGVPTGVQVINGQTLYFDADGRQVKQQRVIGNQRY 1276
QY 1459 GRISYYDANSERVRI 1474
Db 1277 ----WMDKDSGEMKKI 1288
RESULT 9
TJ30857
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30857
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: Z20909; MUID:95122197; PMID:7822030
A:Accession: T30857
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166E; EMBL:L35495; NID:G662378; PII
C:Genetics:
A:Gene: gtfL

Query Match		41.8%;	Score	3238.5;	DB 2;	Length	1449;			
Best Local Similarity		46.0%;	Pred. No.	5.7e-149;						
Matches		702;	Conservative	215;	Mismatches	429;	Indels	179;	Gaps	37;
QY	1	MDKKVRYKLRKVKRWVTVSV-----ASAVMTLTSLGGLVKADSNESKS-----	45							
DB	1	MDKKVHYKHKVKKQWVTIAVGLSLGAVSAVSLGT-NDGVVQADSDHTDATVAIPDITVD	59							
QY	46	--QISNDSNTSVVTANEESNVITEATSKQEAASSQTNHT-----VTTSSSSTSVNPKVYV	99							
DB	60	TGTVSNDT-----TAAQDPTTAAATN--DVATDQATPTATPDLTDTTNTVAANAADTV	112							
QY	100	SNPYTVG-----ETASNGEKLQ--NQTTTVDKTS--EAA-----	129							
DB	113	A---TVGTDRAATNTDNTATNDTAVDTTNNNTTDTTDTTVDRAAATERRATGARRGPTGG	169							
QY	130	--ANNISKOTTEAD--TDVIDDSNAANLQILEKLPNVKEIDGKYVYDNNKQVNTFTL	184							
DB	170	RRATPVNGTNNANNTVTVVNDLPATNNVVDTPGSPHKTNGKQYVVEDDGTIRKNYYL	229							
QY	185	-----IADGKILHFDETGAYTDTSTIDTVNKDITVTRSNLYKKYNQVYDRS	229							
DB	230	BRIGSQYFNAETGELSNOKEYRFDKNGG-TGSSADSTNTNTVNGDK-----NAPYGT	283							
QY	230	AQSEFHVHDHYLTASWYRPKYILKDGKTWTQSTEDKFRPLMTWMPDQBTQRYVNYMNA	289							
DB	284	DKDIELVDGYFTANTWYRPKEILKDGKEWTASTENDKRPLLTVMWPSKAIQASYLYNYKE	343							
QY	290	Q-LGINKTYDDT'SNOQLNIAAATQAKIEAKITLKNYDMLROTISAFVKTSQSAWNSDS	348							
DB	344	QGLGNTQTYTSSQTMQDQAALEYQKRIEERIAREGNTDMLRTTIKNFVKTPQGNMSTS	403							
QY	349	EK-PFDDHLQNGAVLYDNEGKLTVPVANSNYRILNRTPTNOTGKDPRTYADNTIGYEPL	407							
DB	404	ENLDNDNHLQGGALLYNDSR-TSHANSDYALLNRTPTSTQTKHNPKYTKDTSNGGPEPL	462							
QY	408	LANDVDNSNPVQAEQLNLWHLFMFNGNIYANDPDANFDSIRVDVNDVADLLQIAGDY	467							
DB	463	LANDIDNSNPVQAEQLNLWHLFIMNIGTITGSEDENFDFGRVDAVDNVNADLLQIADY	522							
QY	468	LKAAAGIHKNDKAANDHLSILEAKSNDNTPYLHDDGDNMINMDNKLRLSLFLSLAKPLNQ	527							
DB	523	FKAKYGADQSQDQAATKHSILEAWSHNDAYNEDTKGAQLPMDPMHLALVYSLLRPTGN	582							
QY	528	RSQMNPILTSLVNRNTDQNAETAAVPSYFIRAHDSVEQDLADIKAELINPNVVGYSFT	587							
DB	583	RSQVEPLISLNSLNDRESGSKSKEMANYAFVRAHDSVEQSIIGQIKNEIINPOSTGNTFT	642							
QY	588	MEEIKKAFIYNKDLLATEKKYTHYNTALSYALLTNKSSVPRVYVYGMFTDDGQYMAHK	647							
DB	643	LDEMKKAFIYNKDMRSANKQYQYNIPSAVALMLTKDTPRVYVYGMFTDDGQYMAQK	702							
QY	648	TINYEATETLLKARIKYVSGQAMRNOQVGNSE-----LIITSVRYKGKALKATDQD	699							
DB	703	SPYDAIETLLKGRIRYAGGQDMKVNIGYNTNGMDAAGVLTSVRYGTGANSASDGT	762							
QY	700	RTTRTSGVAVIBEGNPNLSRLKASDRVVMGAAHKNQAVRPLLLTDDNGIKAYHSDQAA	759							
DB	763	AETRNOQNAVIVSNQPALRL--TSLNTINMGAAHRNQAVRPELLLTNDGVATYLNDSAN	820							
QY	760	GLVRYTNDRGELIFTAADI KGYANPQVSGYGLGVVPGVGAALIKMFALRLARPHQOMAS--	817							
DB	821	GIVKYTDGNGNLTFANEIRGNPQVDGYLAVVVPVGAS--ENQDVRVAPEKKNSSGL	878							
QY	818	--VHQAALDSRMFEGFSNQAFATKBEYTNVJIAKNDVKFAEAGVTDTEWAPQVYGST	876							
DB	879	VYESNAALDSQVIEYGFNFQFVQNPQSYTNKKIAENANLPKSWGITSFEFAPQYVSSD	938							
QY	877	DGSLFDSVIQNGYATFDRYDGLGSKPNKYGTADDLVKAIKALHSGIKVMADWPDQMYA	936							
DB	939	DGSLFDSVIQNGYATFDRYDGLGSKPNKYGTADDLVKAIKALHSGIKVMADWPDQYIN	998							
QY	937	FPEKEVWVATRVYDKYGPVAGSQIKNTLYVVDGKSSGKDDQQAQKYGCAFLLELOAKYPLF	996							

Db	999	LPGEDEVVATRVNNYGETKDGAIIDHLSYAAKTRTFGNDYQKYGCAFLDELKRLYPQIF	1058
Qy	997	ARQOISTGVPMDFSVKIKOWSAKYFNGTNIILGRGAGYVLKQDQATNTYFNI SDKENKFNFLP	1056
Db	1059	DRVQISTGKRWTTDEKITQWSAKYMTGNTIILDRGSEYVLKN-GLNGYYTGNGKV--SLP	1115
Qy	1057	KTL-----LNQDSQVGFSDYDGKGV-----YYSTSGYQAKNTFISEGD-KWYYFDNN	1102
Db	1116	KVVGSNQSTNGDNQNG---DGSCKFEKRLFSVRYRYNNGQYAKNAFIKNDGNNVYFDNS	1172
Qy	1103	GYWVTGAQSGINGVNYFPLSNGLOLRDAIILKNEDGTAYVYNGDGRYVYQPMGSGVWRH	1162
Db	1173	GRMAVGEKIDGKQYFFLANGVQLRDGYRON-----RR-----	1205
Qy	1163	FNNGEMSVGLTVIDGQVQYFDEM-GYQAGKFTVTADGKIRYFDKOSGNMYNRFIENEE	1221
Db	1206	-----GQVFYDQNGVLNANGK-----QDPKPDNNNNSG---RNQFVQIGN	1244
Qy	1222	GKWLVLGEDGAATVTSQTINGQHLYFRANGVQVKGFEVTDHHRISYYDNGSGDQIRNRF	1281
Db	1245	NWVAYYDNGKRVTHQNINGQLFEDNNGVQVKGRTVNE-NGAIRYYDANSSEMARNRF	1303
Qy	1282	VRNAQGFYFDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDNGSGDQ	1341
Db	1304	AEIEPGVWAYFNNDGTAVKGSQINGQDLYFDQNGRQVKGA-LANVDGNLRYVDVNSGEL	1362
Qy	1342	INRFRVNAQGFYFDNNGYAVTGARTINGQHLYFRANGVQVKGFEVTDHGRISYYDG	1401
Db	1363	YRNRF-HEIDGSWYFFDGNNAVKGMVNINGQLLFDNNGKQIKGHLVR-VNGVRYVDP	1420
Qy	1402	NSGDOIIRNFRVNAQGFYFDNNG	1426
Db	1421	NSGEMAVNRVEVSPGWVYFDGEG	1445
RESULT 10			
T30552			
glucosyltransferase N - Streptococcus salivarius (fragment)			
C;Species: Streptococcus salivarius			
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004			
C;Accession: T30552			
R;Jaffe, R.I.			
submitted to the EMBL Data Library, February 1998			
A;Description: Streptococcus salivarius V1477 gtfN.			
A;Reference number: Z20854			
A;Accession: T30552			
A;Status: preliminary; translated from GB/EMBL/DBJ			
A;Molecule type: DNA			
A;Residues: 1-1449 <JAF>			
A;Cross-references: UNIPROT:O68542; UNIPARC:UPI00000B10FD; EMBL:AF049609; NID:G7			
C;Genetics:			
A;Gene: gtfN			
Query Match			
41.7%; Score 3229.5; DB 2; Length 1449;			
Best Local Similarity			
45.8%; Pred. No. 1.6e-148;			
Matches 698; Conservative 218; Mismatches 430; Indels 179; Gaps 36;			
Qy	1	MDKKVRYKLRKVKRWVTVSV-----ASAVMTLTSLGGLVKADSNESKS-----	45
Db	1	MDKKVHYKHKVKKQWVTIAVGLSLGAVSAVSLGT-NDGVVQADSDHTDATVAIPDITVD	59
Qy	46	--QISNDSNTSVVTANEESNVITEATSKQEAASSQTNHT-----VTTSSSSTSVNPKVYV	99
Db	60	TGTVSNDT-----TAAQDPTTAAATN--DVATDQATPTATPDLTDTTNTVAANAADTV	112
Qy	100	SNPYTVG-----ETASNGEKLQ--NQTTTVDK--TSEAA-----	129
Db	113	A---TVGTDRAATNTDNTATNDTAVDTTNNNTTDTTNTVAATERRATGARRPTCG	169
Qy	130	--ANNISKOTTEAD---TDVIDDSNAANLQILEKLPNVKEIDGKYKYYYDNNKQVNTFTL	184
Db	170	RRATPVNGTNNANNTVTVVNDLPATNNVVDTPGSPHKTNGKQYVVEDDGTIRKNYYL	229

QY 185 -----IADKILHFBETGAYTDTSDTWNKDIIVTRSHLYKKYNQVDRS 229
DB 230 ERIGSQYFNAETGLSNQKEYRFDKNKG-TGSSADSTNTVTVNGDK-----NAFYGT 283
QY 230 AQSPFHVHLYLTAESWYRPKYILKDGKWTQSTEDKPRPLMTWMPDQTOROYVNYNA 289
DB 284 DXDIELVDCYFTANTWYRPKEILKDGKWTASTENDKRPPLTVMWPSKAIQASLYNWK 343
QY 290 Q-LGINKTYDDTSNQLNLNAAATQAKIEAKITTLKNTDMLRQTSIAFVKTSANSDS 348
DB 344 QGLGNTQYTSFSSTQMDQAALVQKRIEGRAREGNTDMLRTTIKNFVKTPQCNSTS 403
QY 349 EK-PFDDHLONGAVLYDNEGLKTPYANSYRILNPTNQTGKDPRTADNTIGYEFL 407
DB 404 ENLNDNDHLQGGALLYNDSR-TSHANSDYRLNRP7SOTGKNPKYTKTDSNGGFBFL 462
QY 408 LANDVDNSPVVQABQLNLHFLMNFNMIYANDPDNFDISRDVADVDNVDADLLQIAGDY 467
DB 463 LANDIDNSNPAVQABQLNLHFLYIMNIGITGSGEDENFDGVKRDVADVDNVDADLLQIAGDY 522
QY 468 LKAAKGIIHKNKAANDHLSILEAWSNDPTPYLHDDGDNMINNDNKLRLSLFSLAKPLNQ 527
DB 523 FKAKYGADQSQDAIKHLSILEAWSHNDAYNEDTKGAQLPMDPMHLALVYSLLRPIGN 582
QY 528 RSGMPLITNSLVNRTDNDNAETAAPVSYSFRAHDSVODLIADIKAIEINPNVVGYSFT 587
DB 583 RSGVEPLISLNSLDRSESQKSKRMANYAFVRAHDSVQSIIGQIKNEINPQSTGNTFT 642
QY 588 MEEIKAFEIYNKOILATYKKYTHNTALSVALLTNTKSSVPRVYVGMFTDDGQYMAHK 647
DB 643 LDEMKAFEIYNKWRANKQYQTNIPSAVALMTHKDTPRVYIGDMYTDGQYMAQK 702
QY 648 TINYBAIETLKARIKYYSGQAMRNQOVENSE-----IITSVRYGKGAKATDGD 699
DB 703 SPYDAIETLLKGRIRYAAAGQDMKVNYIGYNTNGMDAAGVLTSVRYGTGANSADTGT 762
QY 700 RTRTSQVAVTEGNPISRLKASDRVNMGAHQNQAYRPLLLTTDNGIKAYHSDQAA 759
DB 763 AETRNOGMVIVSNQPALRL--TSNLTINMGAHRNQAYRPLLLTTDNGVATYLNDSAN 820
QY 760 GLAVRYTNDRGELIFTAADIKGYANQVSGYGLGVVPGVGAALIKMPALARPHQOMAS-- 817
DB 821 GIVKTDGNGNLTFPSANEIRGNFQVGYLAVVPVGAS--ENQDVRVAPSKENSSGL 878
QY 818 -VHQNAALDSRMVPEGFSNFQAFATKGBEYTNVVIKNDVKRPAEAGVTDTFEMAPQYVSST 876
DB 879 VYESNAALDSQVIYEGFSNFQDFVNPSPQYTNKKIAENANLFPKSWGITSFEPAPQYVSSD 938
QY 877 DGSFLDSVIQNGYATDRYDGLISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQMYA 936
DB 939 DGSFLDSVIQNGYATDRYDGLISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQMYA 998
QY 937 FPEKEVVTATRVKDYGTVPAGSOIKNTLYVVDGKSSGKQQAQYKGAFLLEELQAKYPELF 996
DB 999 LPGDEVTATRVNNYGETDGAIIIDHSIYAAKTRTFGNDYQKYKGAFLDELKRLYPOIF 1058
QY 997 ARKOISTGVPMDPSVKIKQWSAKYPNGTNIILGRGAGYVLKQDQATNTYFNISDNKEINFLP 1056
DB 1059 DRVQISTGKRTWTTDEKTIKWSAKYNGTNIILDRGSEYVLKN-GLNGYGYTNGGKV--SLP 1115
QY 1057 KTL-----LNQDSQVFPYDGKGYV-----YYSTSGYQXQNTFISEGD-KWYTFNN 1102
DB 1116 KVVGSNQNTGNDQNG---DGSQKPEKRLFSVRYRYNNGQYAKNAFIKDNDGNVYFDNS 1172
QY 1103 GYMTGAOSINGVNYFELSNGLQLRDAIILKNEDGTAYYNGDGRYENGYYQFNSGVWRH 1162
DB 1173 GRMAVGEXTIDGKQYFFFLANGVQLRDGYRQN-----RR----- 1205
QY 1163 FNNGMSVGLTVIDGQVQYFDEMGOYQAKGKFTVTTADGK-IRYFDKQSGNMVNRNFIENE 1221
DB 1206 -----GQVFFYDQNG-----VLSANGKQDPKPDNNNTSGRNQFVQIGN 1244

QY 1222 GKWLYLGEDGAAYTGSQTINGOHLIFRANGVQVKGEFVTDHHGRISYYDNGSGDQIRNRP 1281
DB 1245 NWAYYDNGQGRKRVIGHQNINGQELFFDNNGVQVKGRTVNE-NGAIRYDANSGEWARNR 1303
QY 1282 VRNAQGOWFYFDNNGYAVTGARTINGQLLYFRANGVQVKGEFVTDYGRISYYDNGSGDQ 1341
DB 1304 AEIEPGWYFNNBDGTAVKGSQNINGDLYFDQNGRQVKGA-LANVDGNLRYDYVNSGEL 1362
QY 1342 IRNRPVRNAQGOWFYFDNNGYAVTGARTINGOHLIFRANGVQVKGEFVTDHHRGRISYYD 1401
DB 1363 YRNRFP-HEIDGSWYFYDGNNAVKGMVNINGQLLFDNNGKQIKGHLVR-VNGVVRYFDP 1420
QY 1402 NSGDOIRNRFVRNAQGOWFYFDNNG 1426
DB 1421 NSGEMAVNRWVEVSPGWWVYFDGEG 1445
RESULT 11
A44811
glucosyltransferase (EC 2.4.1.1-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A44811; S22726; S28809
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J:Gen. Microbiol. 137, 2577-2593, 1991
A>Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A:Reference number: A44811; MUID:9214837; PMID:1838391
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GIF>
A:Cross-references: UNIPROT:Q00600; UNIPARC:UPI00000BEP31; EMBL:Z11873; NID:g47526; PIDN:44811
A:Note: sequence extracted from NCBI backbone (NCBI:81050, NCBI:81052)
C:Genetics:
A:Gene: gtfJ
C:Keywords: glycosyltransferase; hexosyltransferase
P:1307-1326/Domain: cpl repeat homology <CP4>
Query Match 41.3%; Score 3196.5; DB 2; Length 1518;
Best Local Similarity 45.2%; Pred. No. 6.7e-14;
Matches 695; Conservative 233; Mismatches 472; Indels 139; Gaps 31;
QY 1 MDKVRVYKLRKVKRVTVSVASAVMTLTLTSLGGI-----VKADSNESKQISNDSNTS 54
DB 1 MENKHYKLHKVQKWVITIAS--VALATVGLGSLVTTSSVSADETQDKVTQNSGTT 58
QY 55 -----VVTANEESNVITEATSKQBAASSQTNHVTVTSSSTS 91
DB 59 ASLVTSPKATKADKRTNTEADVLTPAKETNAVETATTTNTQATAEAATTATTADVA 118
QY 92 VNPKEVSVNPYTVGETASNGEKLQNTTVDKTSEAAAANNISKOTTETADTVIDDSNA 151
DB 119 AVPNKEAV---VTTDAPATVTEKABEQPATV---KAEVNVNTEVKAPEA---ALKDS--- 165
QY 152 NLOILEKLPNVEIDGKYVYDNNNGKVRNTFTLIADGKILHFDGTGATVDTSIDTVNKOI 211
DB 166 EVEAALSILKNIKIDGKYVYVNEGSHKENFAITVNGQLLYFGDKCALTSSTSYFTPG- 224
QY 212 VTTRSNLYKYNQVYDRSAQSFHVHDHYLTAEWSYRPKYILKDGKWTQSTEKDPRPLM 271
DB 225 TTNIVDGFSINNRAVDSSEASFELIDGYLTADSWYRPASIIKDGVTWQASTAEDFRPLM 284
QY 272 TWDPDQETQROYVNMNAQLGINKTYDDTSNQLNLNAAATQAKIEAKITTLKNTDMLR 331
DB 285 AWPENVDVTQVNYLNMYSKVFNLDAKYSTDKQETLKVAAKDIQIKIEQIKQAEKSTQMLR 344
QY 332 OTISAFVKTSQAWNSDSEK----PFDDHLONGAVLYDNEGLKTPYANSYRILNPTNQT 387
DB 345 ETISAFVTKQWKNKETENYSKGGEDHLQGGALLYVNDNR-TPWANSDYRLNRTATNQ 403
QY 388 TG-----KQDPRTYADNTIGYEFLANDVDNSNPVQAEOLNLHFLMNFNFIYA 438
DB 404 TGTIDKSLDQSDP-----NHMGDFDPELLANDVLSNPVQAEQLNQIHYLMWNGSIWM 458

QY 921 -----KGIKVMADWV-----PDMYAPPEKEVVVTRV-----D 949
Db 1020 GYIYFDSTGCKAQKGFVTLGKTYFTNTNTMTGFWANNLNLYYFDNEGVMQGTWINVNS 1079
QY 950 KY-----GTPVAGSQIKNTLYVVDGSSKQDQAKYGGAPLELQAKYPFLPARKQISTG 1004
Db 1080 RYFSTAGSVTGQ-----TIDGNKYCFDSNG-----AIYTDVVTGSTYG 1122
QY 1005 VPMPPSVKIKOMSAKYFNGTNILGRGAYVYLKQDQATNTYFNISD-----NKEIN 1053
Db 1123 ENTID-GIMLTGMQTIrYN-----RGY-----SSYFTYFNSDGTAKTGFFYLNKTY 1170
QY 1054 FLPTLLNODSQV-----GFSVDGKGYVYSTSGYQAKNTFISEGDKWYTFDNNNGVMYTGAS 1111
Db 1171 FNP-----SDGRMLQYQVINGNHYYFAPDG-TMOTGMITNGSSKYLLDPGAAVTLQOT 1224
QY 1112 INGVNYFELSNGLDAILKNEGTG---AYGNDGRRYENGYYQFMSGVHRFPNNGEM 1168
Db 1225 INGNKYCFDSNG-----ILOHNGIIFYGNTYYGSD-----NNGIM 1259
QY 1169 SVGLTVIDGOVQYFDEMGOAKGKFTTADGKIRYFDKQS-----GNMYNRRTIENBEGK- 1223
Db 1260 LGLQLQLINGLYCFNSDGSVKTG--LVTYLGTKYTFDSYVSGFQNNNTYFNGNDGTM 1317
QY 1224 ----WL-----YLGEDGAATGSGQTINGQHLRYFRANGVQVKGFEVTDHHRISYYDGS 1273
Db 1318 QTGWNYGYRYRLNDSGIKVTGWTIDGNKYF-----DYYGAKT 1358
QY 1274 GQIIRNFRVNAQGWYFDDNGYAVTGARTINGQLLYFRANGVQVKGFEVTD----- 1325
Db 1359 G-----IVNDIGNYFGNNSGMLTGWOHINGSTYFNSNGIANTG-FITYLGTKYTF 1410
QY 1326 DRYGRI-----SYDNGSG-----DQIRNFRVNA----- 1350
Db 1411 DSYGRWQISMTINGTSYFYANGVMKSTSDSPNTLAVCWWRDSYYQYLYLNAAGTKLTG 1470
QY 1351 ----OQWYFDDNGYAVTGARTINGQHLRYFRANGVQVKG-----E 1387
Db 1471 LQTDIGNTYFDSNGIMQGTIITINGNRYGFGVNGVWLYGLOFINNNTYNSYSGISQNG 1530
QY 1388 FVT-----DRHGR-----SYDNGS-----GDOI 1407
Db 1531 FVTLGNTYFDSYGEIRGLTYINNNYYFNSKGMETGWSYLYRYANPNGLITGFTOT 1590
QY 1408 RN--RFVRNAQGWYF-----FDNNGYAVTGARTINGQHLRYFRANGVQVKGFEV 1454
Db 1591 INKTYFNSDGLLYDQYNGSYGFDKNGVWLYGLOTGICNTYLYNSGISQSG-FI 1649
QY 1455 TDYGRISYYDANSGERVRI 1474
Db 1650 T-LNGKTYFDSYNGRTGI 1668

RESULT 14
A37184
glucan-binding protein - Streptococcus mutans
C:Species: Streptococcus mutans
C>Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 09-Jul-2004
C:Accession: A37184
R:Banas, J.A.; Russell, R.R.B.; Ferretti, J.J.
Infected. Immun. 58, 667-673, 1990
A:Title: Sequence analysis of the gene for the glucan-binding protein of Streptococcus m
A:Reference number: A37184; MUID:90170123; PMID:2307516
A:Accession: A37184
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-563 <BAN>
A:CROSS-references: UNIPARC:UPI000008CB5; GB:M30945; NID:gi53637; PIDN:
F:169-188/Domain: cpl repeat homology <CP1>
F:264-283/Domain: cpl repeat homology <CP2>
F:349-368/Domain: cpl repeat homology <CP3>
F:504-523/Domain: cpl repeat homology <CP4>
F:525-548/Domain: cpl repeat homology <CP5>

Query Match 6.1%; Score 473; DB 2; Length 563;
Best Local Similarity 30.7%; Pred. 1.5e-15;
Matches 139; Conservative 75; Mismatches 145; Indels 94; Gaps 22;
QY 1106 VTG-AQSINGVNYFELSNGLQRLDAILKNEGTGTYAYYNGDGRYENGYYQFMSGVHRHFN 1164
Db 122 VSGEAQSVNAP---SENAAQOETA--KTEPATAA--ENNDAAPTNSFFK-KDGKWKYK 173
QY 1165 -NGEMSVGLTVIDGOVQYFDEMGOAKGKFTTAD-----GKIRYF 1204
Db 174 ADGQLATGWQIIDGKQLYFNQDSQVKGHEIHVETGQIYHPVFISSDPSVLEVNKIYYF 233
QY 1205 DQSGNMVNRRTI-----EN--BEKWLILGEDGAAVVTGSGOTINGQHLRYFRANGV 1252
Db 234 DPDSGELWKDRFVSYSDAPLHYENIKHEG-WFYLGEDGKAALGWTIGGKYYFDTNGV 292
QY 1253 QVKG-----FVTDHHRISYYDNGSGDQIRNFRV-----NAQG-----QWYF 1292
Db 293 QVKGKLIISTDGNVNLISQYKKSFLDPDTGEATNRFVNAKYFYFNAGYVSTTDWFM 352
QY 1293 DNNGYAVTGARTINGQLLYFR-ANGVQVKGFEVTDYGRISYYDNGSGDQIRNFR----- 1346
Db 353 GADGIGVTDWQKIDGMHYFEPSPSGIQVKGD-IAERDGGKYYLDDESDGQVKNRFGTTA 411
QY 1347 -----VRNAQGWYFDDNGYAVTGARTINGQHLRYF-----RANGVQVKGFEVTD 1391
Db 412 ERISTVEARFPKTYFEGADGSRKDLTGWQIIDGKYYFKDDHSIKAKSEYSGIGSVDPD 471
QY 1392 RHGRI-----SYDNGSGDQIRNFRV--NAQGWYFDDNGYAVTGARTINGQHLRYF---- 1442
Db 472 GFAIDGCGYFPDTCQFVTRNFRVKYDYSNIWYVYSGDKRVSGWQTIDGKRYYSQDE 531
QY 1443 RANGVQVKGFEVTDYGRISYYDANSGERVRI 1475
Db 532 KTKGQIRGQITID-GREYTFDKDSGEVINS 563

RESULT 15
A37052
toxin A - Clostridium difficile
C:Species: Clostridium difficile
C>Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 24-Sep-1999
C:Accession: A37052; A60991; S21897; S22437; S08638
R:Dove, C.H.; Wang, S.Z.; Price, S.B.; Phelps, C.J.; Lyster, D.M.; Wilkins, T.D.; Johns
Infected. Immun. 58, 480-488, 1990
A:Title: Molecular characterization of the Clostridium difficile toxin A gene.
A:Reference number: A37052; MUID:90129305; PMID:2105276
A:Accession: A37052
A:Molecule type: DNA
A:Residues: 1-2710 <DOV>
A:CROSS-references: UNIPARC:UPI000016EAD4; GB:M30307; NID:gi44925; PIDN:AAA23283.1; PID
R:Wren, B.W.; Clayton, C.L.; Tabagchali, S.
FEMS Microbiol. Lett. 70, 1-6, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin A gene fragment and detecti
A:Reference number: A60991
A:Accession: A60991
A:Molecule type: DNA
A:Residues: 1-154 <EIC>
A:CROSS-references: UNIPARC:UPI0000178991; EMBL:X60984
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerbo
Mol. Gen. Genet. 233, 260-266, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A:Reference number: S22434; MUID:92293124; PMID:1603068
A:Accession: S22437

Db 2142 GYTIINGKHFYNTDGIQGVF--KGPNGFEYFAPANTDANNIEGQAILYQNEFL-TLN 2198
QY 1222 GKWLYIGEDGAAVTGSQTINGQHLYFRAN-----GQVKGFEVT 1260
Db 2199 GKYYFGSDSKAVTGWRIINNKKYFNPNNAAIAHLCTINNDKYYFSDGILONGYITI 2258
QY 1261 DHGRISYYDGNSGDOI-----RNRFVRNAQGW 1289
Db 2259 ERNN--FYFDANNESKMTVTGVPKGPNGFEYFAPANTHNNNIEGQAILYQNKEL-TINGKK 2315
QY 1290 FYFDNNGAVTGAARTINGQLLYFRANGVQVKGFEVTDYGRISYYDGNSGDOI RNRFVRN 1349
Db 2316 YIFDNDKAVTGWQITIDGKKYFNLNTAEATGWQITID-GKKYFNLNTAEATG--WOT 2372
QY 1350 AQGWYFDNNGY-AVTGAARTINGQHLYFRANGVQVKGFEVTDYGRISYY-----DGN 1403
Db 2373 IDGKKYFNTNTFIASGTYSINGKHFYNTDGIQGVF--KGPNGFEYFAPANTDANN 2430
QY 1404 --GDOI--RNRFVRNAQGWYFDNNGYAVTGAARTINGQHLYFRAN-GVQVKGFEVTDY 1458
Db 2431 IEGQAILYQNKEL-TLINGKKYFGSDSKAVTGLRTIDGKKYFNTNTAVAVTGWQITIN-- 2487
QY 1459 GRISYYDANS 1468
Db 2488 GKYYFNTNT 2497

Search completed: February 11, 2006, 19:41:45
Job time : 40.4446 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 18:57:35 ; Search time 199.028 Seconds
(without alignments)
5228.676 Million cell updates/sec
Title: US-10-797-821-34
Perfect score: 7741
Sequence: 1 MDKVRVYKLRVKRVTVS.....DRYGRISYDANSRVRIN 1475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : UniProt_05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7535.5	97.3	1476	1	GTFB_STRMU
2	5551.5	71.7	1455	1	GTFB_STRMU
3	4582	59.2	1590	2	Q55263_9STRE
4	4580	59.2	1590	2	Q55263_9STRE
5	4578.5	59.1	1597	1	GTF1_STRDO
6	4331	58.5	1592	1	GTF2_STRDO
7	3879.5	50.1	1462	1	GTFD_STRMU
8	3669.5	47.4	1577	2	Q54178_STRGN
9	3650	47.2	1454	2	Q69A94_LEUME
10	3597.5	46.5	1575	2	Q9LCH3_STROR
11	3523.5	45.5	1512	2	Q9WXJ5_9STRE
12	3512.5	45.4	1506	2	Q56CX8_9STRE
13	3506	45.3	1577	2	Q55265_STRSL
14	3434.5	44.4	1554	2	Q8XZL5_9STRE
15	3398.5	43.9	2835	2	Q8G9Q2_LEUME
16	3289	42.5	1599	2	Q00599_STRSL
17	3246	41.9	1290	2	Q48756_LEUME
18	3243	41.9	1508	2	Q52224_LEUME
19	3238.5	41.8	1449	2	Q52264_STRSL
20	3230	41.7	1508	2	Q9EZH5_LEUME
21	3229.5	41.7	1449	2	Q68542_STRSL
22	3215.5	41.5	1477	2	Q91466_LEUME
23	3197	41.3	1527	2	Q8KRRE_LEUME
24	3196.5	41.3	1518	2	Q00600_STRSL
25	3178.5	41.1	1330	2	Q84CN4_LEUME
26	3171.5	41.0	1527	2	Q9ZAR4_LEUME
27	3168.5	40.9	1522	2	Q6TXV4_LEUME
28	3160	40.8	1561	2	Q5SBM8_9LACO
29	2945.5	38.1	1365	1	GTF5_STRDO
30	2897	37.4	1338	2	Q9WXJ4_9STRE
31	2877.5	37.2	1595	2	Q5SBM3_LACSK

32	2793	36.1	1463	2	Q5SBM6_LACFE
33	2537	32.8	2057	2	Q9RE05_LEUME
34	2402	31.0	1016	2	Q9LCJ7_LEUME
35	2343	30.3	1781	2	Q4JCS4_LACRE
36	2340.5	30.2	1772	2	Q5SBM0_LACRE
37	2340	30.2	1772	2	Q5SBM3_LACRE
38	2280	29.5	1781	2	Q4JLC7_LACRE
39	2280	29.5	1781	2	Q5SBL9_LACRE
40	2272	29.4	591	2	Q8VUH3_STRMU
41	1507.5	19.5	1619	2	Q5SBM0_LACRE
42	1493.5	19.3	1231	2	Q5SBM1_LACRE
43	930	12.0	522	2	Q8VV10_STRSA
44	794	10.3	374	2	Q6ZX19_9LACO
45	553	7.1	2817	2	Q97K42_CLOAB

ALIGNMENTS

RESULT 1
GTFB_STRMU STANDARD; PRT: 1476 AA.
AC P08987; O69381; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfB; OrderedLocNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040695;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
play a key role in the development of the dental plaque because of
their ability to adhere to smooth surfaces and mediate the
aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.

FT	VARIANT	150	151	SR -> PK (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	256	256	A -> V (in strain GS-5 and strain MT4467).
FT	VARIANT	425	425	R -> D (in strain MT4251).
FT	VARIANT	519	519	Y -> N (in strain MT4245 and strain MT4251).
FT	VARIANT	538	538	R -> K (in strain MT4245 and strain MT4251).
FT	VARIANT	549	545	Y -> F (in strain MT4245 and strain MT4251).
FT	VARIANT	597	597	N -> D (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
FT	VARIANT	600	600	R -> K (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
FT	VARIANT	601	601	A -> T (in strain GS-5).
FT	VARIANT	614	614	M -> T (in strain MT8148).
FT	VARIANT	727	727	T -> I (in strain GS-5).
FT	VARIANT	734	734	A -> V (in strain MT8148).
FT	VARIANT	964	964	L -> F (in strain MT4239).
FT	VARIANT	1113	1113	N -> Y (in strain MT4239).
FT	VARIANT	1118	1118	A -> T (in strain MT4239).
FT	VARIANT	1204	1204	I -> V (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1208	1208	V -> I (in strain MT8148).
FT	VARIANT	1292	1294	DGH -> NGY (in strain GS-5, strain MT4467 and strain MT8148).
FT	VARIANT	1305	1369	Missing (in strain MT4245).
FT	VARIANT	1326	1326	I -> V (in strain MT8148).
FT	VARIANT	1331	1331	T -> A (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1377	1377	R -> K (in strain MT8148).
FT	VARIANT	1398	1398	V -> I (in strain MT8148).
FT	VARIANT	1424	1424	D -> N (in strain MT4239).
FT	VARIANT	1439	1439	V -> I (in strain MT4239 and strain MT8148).
FT	VARIANT	1444	1444	S -> P (in strain MT8148).
FT	CONFLICT	1337	1455	ORLYPKSGVQAKGELITERKGRIKYKYPNSGNEVRNRYR TSSGNYFPGNDGYALIGHVVEGRVYFDENGVYRYASHD ORNHWDYRDRDFGRGSSAVRFRHSRNGFFDNFRF -> HASILSLMVERLESSLOSKVSVNTMILLIPMKFVIM (in Ref. 1).
SQ	SEQUENCE	1455 AA; 162966 MW; 3CB455A99A4FEC86 CRC64;		
Query Match				71.7%; Score 5551.5; DB 1; Length 1455;
Best Local Similarity				73.6%; Pred. No. 3.4e-250;
Matches 1086; Conservative 112; Mismatches 201; Indels 77; Gaps 16;				
Qy	1	MDKKRYKLRKVKKGVTVSVASAVMTLTLSGGLVKADSNESKQISNDSNTSVVTANE	60	
Db	1	MEKKVRFKLKRVKGVTVSVASAVMTLTLSGSLVKADSTDDROQAVTESQASLVTTSE	60	
Qy	61	-----ESNVITEATS KQBAASQTNHTVTTSSSTSVVNPKEVSVNPYTVGETASN	111	
Db	61	AAKETLTATDTSATSQSPTA--TVTDNVSTNQSTNTTANTANFDVKPTTSE----	114	
Qy	112	GEKLNQTTTVDK-----TSEA-----AANNISKQTTTEAD-----	141	
Db	115	-----QSKTDSKDIATSKAVNRLTATGKFVPANNNTAHSRTVTDKIVPKPKICKLQKP	170	
Qy	142	TDVIDDSNAANLOILEKLPNKVEIDGKYYYDNNCKVRTNFTLIADGKILHFDETGATD	201	
Db	171	SSLSQDDIAA-----LGNVKNRKVNGKYYYKEDGTLOKNVALNINGKTFPFDGTGALS	226	
Qy	202	TSIDTVNKDIVTT-RSNLYKKYNOYVRSAGSFEHVDHYLTAESWYRPKYILKDKGKTWQ	260	
Db	227	N7LPSKKNITNDNTNSFAQNOYVSTDAANFEHVDHYLTAESWYRPKYILKDKGKTWQ	286	
Qy	261	STEKDFRLLMTWPDQETQROYVNMNAQLGINKTYDDTSNQLQNLAAATIAQAKIEAK	320	
Db	287	STEKDFRLLMTWPDQETQROYVNMNAQLGINKTYDDTSNQLQNLAAATIAQAKIEAK	346	
Qy	321	ITTLKNTDNLQRTISAFVKTSQAMNSDSEKFPDDHLQNGAVLYDNEGKLTPTVANSNYRIL	380	

Db	347	ITAEKNTNWLQRTISAFVKTSQAMNSDSEKFPDDHLQKALLYSNNSKLTQSANSYRIL	406	
Qy	381	NRTPTNQTGKDDPRYTADNTTGGYEFLANDVDSNPVVQAEQLNLHFLMNFIGNIYAND	440	
Db	407	NRTPTNQTGKDDPRYTADNTTGGYEFLANDVDSNPVVQAEQLNLHFLMNFIGNIYAND	466	
Qy	441	PDANPDSIRVDADVNDVADLLQIAGDYLKAAGIHKNDKAANDHLSILEASDNDTPYLH	500	
Db	467	PDANPDSIRVDADVNDVADLLQIAGDYLKAAGIHKNDKAANDHLSILEASYNTPYLH	526	
Qy	501	DGDNMINMDNKLRLSLFLSLAKPLNOSPNGNPLITNSLVNRTDDNAETAAPVPSYFIRA	560	
Db	527	DGDNMINMDNKLRLSLFLSLAKPLNQRSGWNPLITNSLVNRTDDNAETAAPVPSYFIRA	586	
Qy	561	HDSEVQDLADI IKAIEINPNVVGYSFTMEEIKKAFEIYNKOLLATEKKYTHYNTALSVAL	620	
Db	587	HDSEVQDLIRNI IRAEINPNVVGYSFTMEEIKKAFEIYNKOLLATEKKYTHYNTALSVAL	646	
Qy	621	LLTNKSSVPRVYVYGDMTDDGOYMAHKTINTEALITLLKARIKYVSGGOAMRNOQVGNSE	680	
Db	647	LLTNKSSVPRVYVYGDMTDDGOYMAHKTINTEALITLLKARIKYVSGGOAMRNOQVGNSE	706	
Qy	681	IITSVRYGKGALEKATDGTDRITRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKNQAYRP	740	
Db	707	IITSVRYGKGALEKATDGTDRITRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKNQAYRP	766	
Qy	741	LLLTNDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGVLGVVVPVGAAL	800	
Db	767	LLLTNDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGVLGVVVPVGAAL	826	
Qy	801	IKMFAIRLRA----RPHQQAASVHQNAALDSRVMFEGFSNFQAFATKKEEYTNVIAKNVDK	857	
Db	827	DQ--DVRVAASTAPSTDGSKSVHQNAALDSRVMFEGFSNFQAFATKKEEYTNVIAKNVDK	884	
Qy	858	FAEKGVTDFENAPQVSVSTSGFLDSVTONGVAFYDRYDLGSLKPNKTYGTADDLKAIIKA	917	
Db	885	FAEKGVTDFENAPQVSVSTSGFLDSVTONGVAFYDRYDLGSLKPNKTYGTADDLKAIIKA	944	
Qy	918	LHSGIKVMADWPDQMYAFPEKEVVTATRVDKYGTVPVAGSQIKNTLYVVDGSSGKDOQ	977	
Db	945	LHSGIKVMADWPDQMYALPEKEVVTATRVDKYGTVPVAGSQIKNTLYVVDGSSGKDOQ	1004	
Qy	978	AKYGAFLEELQAKYPELPFARKQISTGVMPDPSVKIKOWSAKYFNGTILGRGAGYVLKD	1037	
Db	1005	AKYGAFLEELQAKYPELPFARKQISTGVMPDPSVKIKOWSAKYFNGTILGRGAGYVLKD	1064	
Qy	1038	QATNTYFN-1SDNKEINFLPKTLLN-----QDSQVGFSDGKGVYVYSTSGYQAKNTFTS	1091	
Db	1065	QATNTYFSLVSDN---TFLPKSLVNPNNHGTSSSVTGLVFDGKGVYVYSTSGYQAKNAFTS	1121	
Qy	1092	EGDKWYTFDNNGVMTGAQSYNGVYFSLNGLQLRDAILKNEDGTAYAYNGDGRYENG	1151	
Db	1122	LGNWYTFDNNGVMTGAQSYNGVYFSLNGLQLRNALYDNGNKVLSYNGDGRYENG	1181	
Qy	1152	YYQFMSGVWRHFNNGEMSVGLTVTDGQVQYFDEMGYQAKGFVTTADKIRYFDQSGNM	1211	
Db	1182	YYLF--GQWRVYFQNGIMAVGLTRIHGAVQYFDASGFQAKGQFITTADGKRLYFDSDSGNQ	1240	
Qy	1212	YRNFPIENECKWLYLGEDGAATVGTSGQTIINGHLYFRANGVQKGFVTDHGHGRISYDG	1271	
Db	1241	ISNRPVRNKGWELFDHNGVAVTGTFTNGQRLYFKPENGQAKGEFIRADADGHLRYDDP	1300	
Qy	1272	NSGQDIRNRFVRNAQGWYFDNNGVYAVTGARTINGQLLYFRANGVQKGFVTDYGRIR	1331	
Db	1301	NSGNEVRNRFVRNKGWELFDHNGIATVGTIRVNGQRLYFKPSNGVQAKGELLITERKGRIR	1360	
Qy	1332	SYDGNSGDQIRNRFVRNAQGWYFDNNGVYAVTGARTINGOHLLYFRANGVQKGFVTD	1391	
Db	1361	KYDPSNGNEVRNRYVTRTSSGNYWYFNGDYGALLIGHVVEGRVYFDENGVRVYASHDQR	1420	
Qy	1392	RHGRISY----YDGNSGDQIRNRFVRNAQGWYFDN	1424	

```
Db 1421 NHWDYDRRDFGRGSSAVFRHSRNG-----FFDN 1451
RESULT 3
Q55263 9STRE
ID Q55263_9STRE PRELIMINARY; PRT; 1590 AA..
AC Q55263;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GTF-1.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=13110;
RN [1]_TaxID=13110;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucon synthetase
RT produced from Streptococcus sobrinus ATCC 33478."
RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).
RN [2]_
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases."
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D63570; BAA09792.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; 1GVM.
DR GO; GO:0009250; P-glucon biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;

Query Match 59.2%; Score 4582; DB 2; Length 1590;
Best Local Similarity 56.5%; Pred. No. 6.2e-205;
Matches 898; Conservative 212; Mismatches 336; Indels 144; Gaps 19;

QY 1 MDKVRKYLKRYKRWVTVSVASAVMTLTTLSSGLVKADSNESKQISNDSTSVVTANE 60
DB 1 MEKNVRFKHKVKKRWVTVSVASATMLASALGASVASADTDTA-----SDDSNQAVVTGDQ 56
QY 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKE-----VVSNPYTVGTASNGEKL- 115
DB 57 TTN--NQATDQTSIAATSEQASSTDAATDQASAEQTQGTASTDTAAQTITNANEAK 114
QY 116 -----NQNTTVKTSAAANNISKQTEADTVIDDSNAANLQILEKLPNVKIDGKY 169
DB 115 WYPTENENOGFTDEMLAE--AKNVA--TAESDSPSPD-----LAKMSNVKQVDGKY 161
QY 170 YYDNNQKRVNTFTLLADGKILHFDGTGAYTDTSDTVNK--DIYTTSLNLYKKYNOYVD 227
DB 162 YYDQDGNVKNFVAVSGDKIYYPDETGAYKDTSKVDADKSSVASQNAIFAAFNRAYS 221
QY 228 RSAQSFEHVDHYLTAEVSRPKYILKDGKTWTQSTKDFRPLMTWMPDOETORQVNYM 287
DB 222 TSAENFEADVNLTDADSVRPKSILKDGKTWETSGKDFRPLMAWNPDTETKRYNYM 281
QY 288 NALQINKYTDTSNQLQNTAAATQAKIEAKITLTKNTDNLROTISAFVKTSQAWNSD 347
DB 282 NLVVGIDKTYTAETSQADLTAAELVQARIEQKITTEQNTKWLREAISAFVKTPQWNGE 341
QY 348 SEKPDDHLONGAVLYDNEGKLTTPVANSYRILNPTPTNQTGKDPRT--ADNTIGGYE 405
DB 342 SEKPYDDHLONGALFKDQSDLTPTDQTSQYRLLNRTPTNQTGSLDSRFTYANDPLGGYE 401
```



```

Db 1296 TGAQTIRGKLYFKANGQQVKGDIVKGTGDKIRYDAKSGEQVFNKTVKAADGKTYVIGN 1355
Qy 1360 NGYAV-----TGARTI 1370
Db 1356 DGVAVDPSVVGKTFKDGASGARFFYNLXGQLVTGSGWYETANHDWYIYOSGKALTGEQII 1415
Qy 1371 NGQHLYFRANGVOVKGFEVTDHRGRISYYDGSNGDQIRNRFV-----1412
Db 1416 NGQHLYFKEDGHQVKGQLVTGTDGKVRYYDANSQDQAFNKSVTVNGKTYFYFGNDGTAQTA 1475
Qy 1413 -----RNAQGMQFYFDNNGYAVTGARTINGOHLFY 1442
Db 1476 GNPKGTFKDGSDIRFYSMEGQLVTGSGWYENAAQQQWLYV-KNGKVLTLGLQTVGSORVYF 1534
Qy 1443 RANGVOVKGFEVTDYGRISYYDANSGERV 1472
Db 1535 DENGIOAKGKAVRTSDGKIRYFDENSGSMI 1564

RESULT 5
ID GTFL_STRDO STANDARD; PRT; 1597 AA.
AC P11061.
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfI;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Galpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus sobrinus MFE28";
RL J. Bacteriol. 169:4271-4278 (1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 19 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.
CC
CC EMBL; M17391; AAC63063.1; -; Genomic DNA.
DR InterPro; IPR002479; Cell wall bd put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1597 Glucosyltransferase-I.
FT REPEAT 1099 1132 A repeat.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1227 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.

```

```

FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 B repeat.
FT REPEAT 1519 1568 AC repeat.
FT REPEAT 1582 1597 A repeat (incomplete).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1.25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MM; B9E86A200868798E CRC64;

Query Match 59.1%; Score 4578.5; DB 1; Length 1597;
Best Local Similarity 55.9%; Pred. No. 9.1e-205;
Matches 892; Conservative 216; Mismatches 339; Indels 149; Gaps 18;

Qy 1 MDKVRVYKLRVKRWVTVSASAVMTLTLTSGGLVKADSNESKSOISNDSNTSVVTAN- 59
Db 1 MEKNERFMVKRWVTISVASATWLASALCASVASADT-----ETVSSDSNQAVLTAQ 56

Qy 60 -----BESNVITEATSKQEAASQTNRHTVTTSSTSVVNPKEVSNPYTVGETASNG 112
Db 57 TTTNQDTQTSVAATATSEQSASTDAATDQASATDQASAAEQTGTTASTDTAAQTITNA 116

Qy 113 EKL-----NQTTTVDKTSAAANNISKQTTEADTVIDDSNAANLQILEKLPVVKXI 165
Db 117 NEAKVVPPTENENQVFTDEMLAEA-----KNVATAESNSIPSD-----LAKMSNVKQV 163

Qy 166 DGKYYYDNGNGKVRTNFTLIADGKILHPEDETCAITDTS---IDTVNKDITVTRSNLYKKY 222
Db 164 DGKYYYDODGNVKNKFNFAVSUGEKIYFDETGAYKDTSKVEADKSGSDI-SKEETTFAAN 222

Qy 223 NQVYDRSAQSEPHVDHYLTAEWSYRPKYILKDGKTTQSTETKDFRPLMTWMPDOSTQR 282
Db 223 NRAYSTSAENFEAIDNYLTADSWYRPKSILKDGKTTWESSKDDFRPLMAWMPDTETKEN 282

Qy 283 YVYNVAOLGINKTYDDTSNQLNLNIAAATQAKTEAKITTKNTDNLQTTISAFVKTQS 342
Db 283 YVYNMKNVVGIDKTYTAETSQADLTAAAEVLQVARIQKITTQNTKWLREASISAFVKTP 342

Qy 343 AWDSEKPEDDHLONGAVLVYDNECKLTPYANSYRILNRTNPTNQTGKDPYV--ADNT 400
Db 343 QNGSESEKPYDDHLQNGALKFNDQSDLTPTQSNYRLLNRTNPTNQTGSLDSRFTYANDP 402

Qy 401 IGGYFLLANDVDSNPVVOAEQLNWLHFLMFGNIYANDPDANFDSIRVDADVNDADL 460
Db 403 LGGYELLANDVDSNPVVOAEQLNWLHFLMFGNIYAKADANFDSIRVDADVNDADL 462

Qy 461 LOIAGDYKAAKGIHKNDAKANDHLSILEASDNDPTPYLHDDGDMNMDNKLRLSLFSS 520
Db 463 LQISSDYLKAAAYGIDGNKNKNANNHVSIVEAWSDDNTPYLHDDGDMNMDNKFRLSMLWS 522

Qy 521 LAKPLNORSGMNPILITNSLVNRTDQNAETAAVPSYSPFIRAHDSVQDILTADIKAIEIN 580
Db 523 LAKPLDKKSGLNPLIHNSLVNREVDRETVPSYSFARAHDSVQDILIRDIKAIEINPN 582

Qy 581 VVGYSFTMEEEKKAFPEIYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVYVYGMFTDD 640
Db 583 AFGYSFTQDEIDQAFKIYNEDLKKTKYTHYNVPLSYLLLTNKGSIPIRVYVYGMFTDD 642

Qy 641 GOYMAHKTINYEATETLKARIKYVSGOAMRNQOVGNSEIITSVRYGKALKATDGTGR 700
Db 643 GOYMANKTVNYDAIESLULKARMKYVAGQAMQNYQIGNGEILTSVRYGKALKQSDKGDA 702

Qy 701 TTRTSGVAVIEGNPNPSLKLKASDRVVVMGAAHKNQAYRPLLLTTDNGTKAYHSDQEA-- 758
Db 703 TTRTSGVGVVMGNQPNFSLDGK-VVALNMGAAHQAQYRALMVSTKDGQVATYATDADASK 761

Qy 759 AGLVRYTNDRCGELIFTAADIKGYANPQVSGYLVGMVVPVGA--LIKNFALRLARPHQOM 815
Db 762 AGLVKRTDENGILYFLNDDLGKVANPQVSGFLQVWVPVGAADDQDIRVAASDTASTDGK- 820

Qy 816 ASVHQNAALDSRVNMFEGSFNFQAFATKKEEYTNVVIKNNVKAFAEAVGVTDFEMAPQYVSS 875
Db 821 -SLHQDAAMDSDRVNMFEGSFNFQSFATKKEEYTNVVIANNVDFKFSWGITDFEMAPQYVSS 879

```

```

QY 876 TDGSLDVSIVNGYAFDREYDLIGSKPNKYGTADDLVKAIKALHSGKIGKVMADWPDQWY 935
Db 880 TDGQLDVSIVNGYAFDREYDLIGSKPNKYGTADDLVKAIKALHSGKIGKVMADWPDQWY 939
QY 936 AFPEKVEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKDQQAQKGGAFLEBLQAKYDEL 995
Db 940 TFPKQEVVTVTFDFKGFPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFLEBLQAKYDEL 999
QY 996 PARKOISTGVPMDSVKIKQMSAKYFNGNINILGRGAGYVLKQDQATNTYFNISDNKEINPL 1055
Db 1000 FTKKQISTGQAITDPSVKIKQMSAKYFNGNINILGRGAGYVLKQDQATNTYFNISDNKEINPL 1057
QY 1056 PKTLLNODSQVGSVDGKGYV-YETSGVQAKNTFISEGDKWYFYDNNNGYMTGQAOSNG 1114
Db 1058 PKTLLGQVVSIGREDFGTYVNSSTTEKVTDSFITEAGNLYIFYGQDGYMTGQAQNLKG 1117
QY 1115 VNYFELSNGLQRLDAILKNEGCTYAYYGNDRYENGYYQFMSGYVRHFNNGEMSVGLTV 1174
Db 1118 SNYFELANGAALNTVYTTDAQOQHYYGNDGKRYENGYYQFMSGYVRHFNNGEMSVGLTV 1177
QY 1175 IDGQVQVDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNRIENEBGKWLVLGBDGAAY 1234
Db 1178 VDGHVQVYEDKQGVQAKDKIIVTRDGKVRYPFDQHGNAVTNTFVADKTGHWYLYLKGXGVAV 1237
QY 1235 TGSOTINGOHLVFRANGVQVKGFEVTDHGRISYYDGNSGDQIARNRFVRNAQOQWYFDN 1294
Db 1238 TGAQTVGQOHLVFEANGQVQKGFVTAKDGLFYDVSQDMMTNTFIEDKAGNWFYLGK 1297
QY 1295 NGYAVTGARTINGQLLYFRANGVQVKGFEVTDHGRISYYDGNSGDQIARNRFVRNAQOQWY 1354
Db 1298 DGAATVGAQTIKGKLYFKANGQVQKGDIVKADGKIRYDAQTGEQVFNKSV-SVNGKT 1356
QY 1355 FYFDNNGVA-----VT 1365
Db 1357 YFYGSDGTAQTOANPKGTFFKQDGLVRYFNLEGQVYSGSGWYETAHEHWVYVKGKVLIT 1416
QY 1366 GARTINGOHLVFRANGVQVKGFEVTDHGRISYYDGNSGDQIARNRFV----- 1412
Db 1417 GAQTIGNRVYFPKONGHVQKQLVGTGNDGKRLRYDANSQDQAFNKSVTNGKTYFYGSBDG 1476
QY 1413 -----RNAQOQWYFDNNGYAVTGARTIN 1436
Db 1477 TAQTOANPKGTFFKQDGLVRYFNLEGQVYSGSGWYKNAQOQWLV-KDGKVLTLGLQVIG 1535
QY 1437 GOHLVFRANGVQVKGFEVTDHGRISYYDANSGERV 1472
Db 1536 NQKVYFDKNGIQAKGKAVRTSDGKVRYPDENGSGMI 1571

RESULT 6
GTF2_STRDO
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (SC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_taxid=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=6715 / Serotype G;
RX MEDLINE=9112327; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996 (1991).

```

```

CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 16 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; D90213; BAAL4241.1; -; Genomic DNA.
CC HSSP; P06653; IGWM.
CC InterPro; IPR002479; Cell wall bd put.
CC InterPro; IPR003318; Glyco hydro_70.
CC Pfam; PF01473; CW binding 1; 3.
CC Pfam; PF02324; Glyco hydro_70; 1.
CC Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
CC SIGNAL 1 38 Potential.
CC CHAIN 39 1592 Glucosyltransferase-I.
CC REPEAT 1093 1142 1.
CC REPEAT 1158 1207 2.
CC REPEAT 1222 1272 3.
CC REPEAT 1287 1337 4.
CC REPEAT 1402 1451 5.
CC REPEAT 1514 1563 6.
CC REPEAT 1577 1592 7 (incomplete).
CC REGION 39 1044 Catalytic (approximate).
CC REGION 1093 1592 7 X tandem repeats.
CC REGION 1093 1592 Glucan-binding (approximate).
CC SEQUENCE 1592 AA; 176168 MW; BC0A66D079351ECP CRC64;
SQ
Query Match 58.5%; Score 4531; DB 1; Length 1592;
Best Local Similarity 55.8%; Pred. No. 1.5e-202;
Matches 889; Conservative 213; Mismatches 344; Indels 146; Gaps 19;
QY 1 MDKVRVYKLRVKRWRVTVSVASAVMTLTITSGGLVKADSNESKSIQISNDSWTSVTVANE 60
Db 1 MEKNVRPKMVKRWRVTVSVASAVMTLTITSGGLVKADSNESKSIQISNDSWTSVTVANE 56
QY 61 ESNVITEATSKQEAASSQTNHTVITSSSSTSVVNPKE-----VSNPYTVGETASNGEKL- 115
Db 57 TTN--NQATDQTSIAATSEQASASTDAATQASAAEQTGTTASTDTAAQTITTNANEAK 114
QY 116 -----NQTTTVDKTTSEAAANNISKQTTADTDVDDSNAAANLQILEKLPNKEIDGKY 169
Db 115 WPTFENENQGTDEMLAEA-----KNVATAESDISIPSD-----LAKMSNVKQVDGKY 161
QY 170 YYDNNGKVRNFTLIADGKILHFDGTGAYTDTSIDTVNK--DIVTTRSNLYKKYNQVYD 227
Db 162 YYYDQDGNVKKNFAVSGDKIYYFDGTGAYKDTSKVDADKSSSAVSQNTAPIAANNRAYS 221
QY 228 RSAQSFEHVDHYLTAEHMYRKYILKDGKTTWQSTEKDFRPLMTWMPDQETQROYVNYM 287
Db 222 TSAKNFEADVNYLTADSWYRPRKSIKDGKTTWQSTEKDFRPLMTWMPDQETQROYVNYM 281
QY 288 NAQLGINKTYDDTNSQLNLNAAATIQAKIEBAKITTLKNTDMLRTIQISAFVKTSQANSD 347
Db 282 NKVYGIDKTYTAETSOADLTAAELVQARIQKITSNNNTKWLREAIQISAFVKTSQANSD 341
QY 348 SEKPFDDHLQNGAVLYDNEGKLTVPYANSNYRIINRTPNQTKGKDPRT--ADNTIGGYE 405
Db 342 SEKPYDDHLQNGALLFDNQTLTPTDQSNRYLLNRTPTNQTKGSLDRFTYFNPNDPLGGYD 401

```


QY 406 FLLANDVNSNPVQAEQNLWHLFLMNFQNIYANDPDANPDSIRVDANVNDVADLLQIAG 465
DB 402 FLLANDVNSNPVQAEQNLWHLFLMNFQNIYANDPDANPDSIRVDANVNDVADLLQIAG 461
QY 466 DYLLAAKGIHKNDKAANDHLSLEAASNDPTVYLDHDDGNMNMNDKRLSLFLPSLAKPL 525
DB 462 DYLLAAKGIHKNDKAANDHLSLEAASNDPTVYLDHDDGNMNMNDKRLSLFLPSLAKPL 521
QY 526 NORSGMNPITNSLVNRDNDNAETAAPVPSYFIRAHSDSEVDLIADIIKAEINPNVVGYS 585
DB 522 DVRSGLNPLHNSLVNDREVDREVEVPSPYSPARAHSDSEVDLIADIIKAEINPNVVGYS 581
QY 586 FTWEEIKKAFIYNKDLLATEKYTHYNTALSYALLTNKSSVPRVYGMFTDDCQYMA 645
DB 582 FTQEEIDQAFKIYNEDLKSDKKYTHYNPLSYLLTNKSGSPRVYGMFTDDCQYMA 641
QY 646 HKTINYEAJETLLKARIKYVSGQAMRNOQVGNSEIITSRYVGKALKATDGTGRTTRS 705
DB 642 NKTIVNYDAIESLLKAMKTVAGQAMQNYQINGEILTSRYVGKALKATDGTGRTTRS 701
QY 706 GVAVIEGNNPSRLKASDRVVMGAAHKNQAVRPLLLTTDNGIKAYHSDQEA--AGLYR 763
DB 702 GGVVMGQNPFSLDGK-VVALNMGAHAHQEYRALMVSTKQGVATYATDADASKAGLYK 760
QY 764 YTNDRBELIFTAADIKGYANPOVSGYLGVVPGVAA---LIXMFAURLARPHQOMASVHQ 820
DB 761 RTDENGILYFLNDLKGAVNPQVSGFLQVVPVGAADDQDQIRVAASDTASTDGK--SLHQ 818
QY 821 NAALSRVMFEFSNFQAFATKEEYTNVIAKNDKFAENGVTDPFEMAPQVSSYDGSF 880
DB 819 DAAMDSRVNFEFSNFQAFATKEEYTNVIAKNDKFAENGVTDPFEMAPQVSSYDGSF 878
QY 881 LDSVLQNGYAFTRDYLGIKSPENKYGTADLVKAIKALHSKGIKVMADWPQOMTAFBPK 940
DB 879 QDSVLQNGYAFTRDYLGSANKYGTADLVKAIKALHAKGLKVMADWPQOMTAFBPK 938
QY 941 EVVTRVDKYGTPVAGSGIKNLTVYVVGKSGKQQAQYGAFLFELQAKYPELPARKQ 1000
DB 939 EVVTRVTRDKFKPTAGSQINHSLYTDTKSSGDDYQAKYGAFLDELKEKPELPFTKQ 998
QY 1001 ISTGVPMDSVKIKQWAKYFNGTILGRGAGVYLKQATNTYFNISDNKEINFLPKTLL 1060
DB 999 METGQAIDPSVKIKQWAKYFNGSILGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1056
QY 1061 NQDSQVGSYDGKGVYVYST-SGYQAKNTFISEGDKWYFDNNGVYMTGAQSGVNYVF 1119
DB 1057 GKVVESGIRYDGKGVYVYST-SGYQAKNTFISEGDKWYMTGAQSGVNYVF 1116
QY 1120 LNSGLQLDAILKNEGDTYAYYNGDGRY--ENGYQFMGVSVRHFNNGEMSVGLTVIDG 1177
DB 1117 LENGTAIRNTIYVTDAGNSHYIYANDGKRYENENGYQF--GNDWRYFKDGNMAVGLTTVDG 1175
QY 1178 QVQYFDEMGYQAKGFVTTADGKIYFDKQSGNMVYRNRIENEGKWLVLGBDGAATGS 1237
DB 1176 NVQYFDPKQGVQAKKIIVTRDGKRYVFDQHGNAVNTFTADKTGHYYLKGDKGAVTGA 1235
QY 1238 QTINGQHLVFRANGVQKGEFVTDHGRISYYDNGSGDQIRNRFVNAQGMQFYFDNGY 1297
DB 1236 QTVGQKLYFANGQVQKGFVTSDEGKLYFVDSGDMWMTFTIEDKAGNFFYLKGQA 1295
QY 1298 AVTGARTINGQLLYFRANGVQKGEFVTDYGRISYYDNGSGDQIRNRFVNAQGMQFYF 1357
DB 1296 AVTGAQTIRGQKLYFKANGQVQKGDIVKGTGDKIRYDAKSGEQVFNKTVKAADGKTYI 1355
QY 1358 DNGYAV-----TGAR 1368
DB 1356 GNDGVAVDPFVVKGTQFKDASGALRFYNLKGLVTSQGWYETANHDWVYIQSGKALTGRQ 1415
QY 1369 TINGQHLVFRANGVQKGEFVTDHGRISYYDNGSGDQIRNRFV----- 1412
DB 1416 TINGQHLVFKDGHQVKGQLVGTGDKVRYIDANSQDQAFNKSVTNGKTYTFGNDGTAQ 1475
QY 1413 -----RNAQGMQFYFDNGGYAVTGARTINGQHL 1440

DB 1476 TAGNPKGTQFKDGSDFRIFYSMEGQLVTGSGWYSNAQGMQLYV-KNGKVLTLGLQTVGSQRV 1534
QY 1441 YFRANGVQKGEFVTDYGRISYYDANSGERV 1472
DB 1535 YFDENGIOAKGKAVRTSDGKIRYFDENSGSMI 1566
RESULT 7
GTFD_STRMU
ID_GTFD_STRMU STANDARD; PRT: 1462 AA
AC P49331; O69389; O69389; O69389; O69392; O69398;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gTfD; OrderedLocNames=SMU.910;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RX [1]_TaxID=1309;
RV NUCLEOTIDE SEQUENCE.
RP STRAIN=GS-5;
RC MEDLINE=91100958; PubMed=2148600;
RA Honda O., Kato C., Kuramitsu H.K.;
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
RT the glucosyltransferase-S enzyme.";
RL J. Gen. Microbiol. 136:2099-2105(1990).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=982331843; PubMed=9570124;
RA Fujimura T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Prineaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-1 synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 6 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; M29296; AAA26895.1; -; Genomic DNA.
DR EMBL; D88653; BAA26103.1; -; Genomic DNA.

	SIGNAL	1	? CHAIN ?	Potential. Glucosyltransferase-S.
R	EMBL; D8656; BAA26107.1; -; Genomic DNA.			
R	EMBL; D8659; BAA26111.1; -; Genomic DNA.			
R	EMBL; D8662; BAA26115.1; -; Genomic DNA.			
R	EMBL; D8957; BAA26121.1; -; Genomic DNA.			
R	EMBL; AE014932; AAN58619.1; -; Genomic DNA.			
R	HSP; P06653; IGM.			
R	InterPro; IPR002479; Cell wall_bd_put.			
R	InterPro; IPR003318; Glyco_hydro_70.			
R	Pfam; PF01473; CW binding I; 4.			
R	Pfam; PF02324; Glyc_hydro_70; 1.			
M	Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;			
N	Transferase.			
T	VARIANT	10	10	3 X 63 AA approximate tandem repeats.
T	VARIANT	19	19	Y --> H (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
T	VARIANT	58	58	I --> V (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT9148).
T	VARIANT	68	68	K --> E (in strain MT4467).
T	VARIANT	81	81	A --> S (in strain MT4239 and strain MT4245).
T	VARIANT	113	113	A --> T (in strain MT4251 and strain MT8148).
T	VARIANT	122	122	T --> I (in strain MT4239 and strain MT4245).
T	VARIANT	132	132	A --> V (in strain MT4239, strain MT4245 and strain MT8148).
T	VARIANT	135	135	A --> S (in strain GS-5 and strain MT4467).
T	VARIANT	137	137	A --> T (in strain MT4245).
T	VARIANT	202	202	A --> T (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
T	VARIANT	255	255	V --> L (in strain MT4239).
T	VARIANT	275	275	D --> N (in strain MT9148).
T	VARIANT	288	288	E --> D (in strain MT4239, strain MT4245 and strain MT4251).
T	VARIANT	301	301	D --> N (in strain MT4239, strain MT4245 and strain MT4251).
T	VARIANT	313	313	Q --> H (in strain MT4245).
T	VARIANT	317	317	D --> N (in strain MT4239 and strain MT4251).
T	VARIANT	328	328	Q --> K (in strain MT4239).
T	VARIANT	350	350	F --> F (in strain MT4239).
T	VARIANT	628	633	F --> L (in strain MT4239, strain MT4251 and strain MT4467).
T	VARIANT	688	688	KKKYIQ --> EEKYYL (in strain MT4251).
T	VARIANT	726	732	A --> S (in strain MT4239).
T	VARIANT	726	730	TDQGSA --> ADKGNDG (in strain MT4251).
T	VARIANT	762	762	TDQGS --> ADKGN (in strain MT4239 and strain MT4245).
T	VARIANT	964	964	T --> A (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
T	VARIANT	1019	1019	D --> Y (in strain MT4251).
T	VARIANT	1059	1060	E --> K (in strain MT4245 and strain MT4251).
T	VARIANT	1060	1060	LG --> IR (in strain MT4251).
T	VARIANT	1080	1080	G --> R (in strain MT4245).
T	VARIANT	1142	1142	G --> R (in strain MT4239).
T	VARIANT	1198	1198	H --> Q (in strain GS-5).
T	VARIANT	1220	1220	S --> N (in strain MT4239).
T	VARIANT	1280	1280	Y --> C (in strain MT4239 and strain MT4467).
T	VARIANT	1282	1282	F --> L (in strain MT4467).
T	VARIANT	1290	1290	Q --> P (in strain MT4245).
T	VARIANT			K --> T (in strain MT4245).

[illegible]

QY 862 GVTDFEMAPQYVSSDGSFLDSVQNGYAFTRDYDLGISKPNKYGPADDLVKAIALHSHK 921
Db 894 GVTSEMAPQYVSSDGSFLDSIIONGYAFEDYDLGILANSKNNKYGSQQDMINAVKALHKS 953
QY 922 GIKWADWPDOMYAPPEKEVVTATRVKGYTPVAGSOIKNTLYVVDGSSGKDOQAKYG 981
Db 954 GLQVADWPQYVSSDGSFLDSIIONGYAFEDYDLGILANSKNNKYGSQQDMINAVKALHKS 1013
QY 982 GAFLELOAKYFELFARKQISTGVPMDPSVKIKOMSAKIFNGTNIILGRGAGVYVLQDQATN 1041
Db 1014 GAFLELOAKYFELFARKQISTGVPMDPSVKIKOMSAKIFNGTNIILGRGAGVYVLQDQATN 1073
QY 1042 TYFNISDNKEINFLPKTLNQDSQVGSFVDGKGYVYVSTSGYQAKWTFISEG-DKWYYPD 1100
Db 1074 KYFELKGNQ--TYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYYPD 1131
QY 1101 NGYVMTGAQSTNGVNYFSLNGLQRLDAILKNEGTYAYYNDGRRYENGYYQF-MSGV 1159
Db 1132 NGHVMYGLQHLNGEVQYFSLNGVQLRESFLENADGSKNYFGLHGNRYSGYSEFNDGSK 1191
QY 1160 MRHFN-NGEMSVGLTVIDQVOYFDEMGYQAKGKFTTADGKIRYFDKQSGNNYRNRFT 1218
Db 1192 WRYFDASGVMAVGLKTINGTQYFDQDGYQVKGAMITGSGKRYFDGSGNNYRNRFPAN 1251
QY 1219 NEEGKWLILGEDGAAVTGSQTINGHLYFRANGVQVKGFEVTDHHRISYYDNGSDQIR 1278
Db 1252 DKNGDWWYLLNSDIALVQVQTINGKTYFYQDQKQIKGKIITD-NGKLYFLANGSELAR 1310
QY 1279 NREVRNAQOWEYFDNNGYAVGARTINGQLLYFRANGVQVKGFEVTDHHRISYYDNGS 1338
Db 1311 NIFATDSQNNWYFSGDGVAVTGSQTIAKGLYPASDGKQVKGSEFVTYN-GKVHYHHADS 1369
QY 1339 GQOIRNRFRVNAQOWEYFDNNGYAVGARTINGHLYFRANGVQVKGFEVTDHHRISYY 1398
Db 1370 GELQVNRFEADKGNWYLLDSNGEALTGSQRINGQVFTTREGKQVKGVDVADRGLLRY 1429
QY 1399 YDNGSDQIRNRFRVNAQOWEYFDNNGYA 1428
Db 1430 YDKNSGNNYKVVTLANGRRIGIDRWGIA 1459

RESULT 8

Q54178 STRGN PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfG;
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
phase variants.";
RL Dev. Biol. Stand. 85:309-314 (1995).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RA Minick P., Vickerman M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U12643; AAC43483.1; -, Genomic_DNA.
DR PIR; B41898; B41898.
DR HSP; P06653; 1HCK.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.

DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;
Query Match 47.4%; Score 3669.5; DB 2; Length 1577;
Best Local Similarity 48.5%; Pred. No. 2.2e-162;
Matches 759; Conservative 241; Mismatches 459; Indels 105; Gaps 24;
QY 1 MDKRYRYKLRKVKKGVVTVSVASAVMTLT--LSGGLVKADS-----NESKQISNDN 52
Db 2 MEKKVHYKMKVKKQWVAIAVTSLLALVAPKALGESGVYADANQVTVNVKEQSAQSK 61
QY 53 TSVVTANBESNVITEATSKQEAASQTNHTVTSSST----SVNPKEV----- 98
Db 62 DSEQTTSKATDSSQLEVKQASSKEYQASAAANPTANEQTQODKEVEFRTDSRHE 121
QY 99 ----VSNPYVTGEBTASNGEKLQNTTVDKTSSEAANNISKQTTTEADTVIDDSNAANLQ 154
Db 122 LTQKTSDDSEKSGSQEPKVADQAESTDKTQAAQL--AKQDSRANDQEBETENVAKAT 178
QY 155 ILEK-----LPNVKEIDGKYYY 172
Db 179 VSDKIATPKPERLPEPAQRKESITEKMLAAQAAPVNTHEHDDVLAHIKTIDGKYYV 238
QY 173 DNNGKVRTNFTLIADGKILHPD-ETGAYTDTSDITVNDIVTTRSNLYKKNVQVDRSAQ 231
Db 239 QDGVTKKQKPAVELNGKILYFPAETGALVDSN-EYQFQGTSSLNNEFTQKAFYTTDK 297
QY 232 SEFVHDYLTAEWYRPKYLKDGKWTOSTEKDFRPLMLTWPPDQETQRYVNYMNAQ- 290
Db 298 DIETVDGYLTADSWYRPKILKDGKWTASTETDLRPLLMWWPDKRTQINLYNMYNQEN 357
QY 291 LGINKTYDDTSNQLQNTAAATQIAKIEAKITLKNQDWLRTQISAFVKTQSAWN----S 346
Db 358 LGIG-AFESKTEQVLLTNAVQVQRKIEERISKEGDKTLRLTMSAFVKTQPNWNIKTES 416
QY 347 DSEKPFDDHLQNGAVLYDNEGKLPYANSNYRLNARTPTNQTGKDPRTVADNTTGGYEF 406
Db 417 ETGTGKDHQGGALTYNSDK-TSHANSRYRLNRTPTSGT--TPKTFIDKSNNGYEF 473
QY 407 LLANDVDNSNPVQAEQLNWLHFLNFGNIYANDPDANFDSIRVDAVDNVDLQLIAGD 466
Db 474 LLANDVDNSNPVQAEQLNWLHFLNFGNIYANDPDANFDSIRVDAVDNVDLQLIAGD 533
QY 467 YLKAAGTHKNDKAANDHLSILEAWSDNDTPYLHDDGDNMNMNKLRLSLFLAKPLN 526
Db 534 YPKSRYKVGESEEEALKHLSILEAWSDNDPDYNTKDTKGAQLAIDNKLRLSLFLSFMRLKS 593
QY 527 ORSGMNPILITNSLVNRTDNDNAETAAPVSPSIFRAHDSVODLIADI IKAELNPVNVGYSP 586
Db 594 IRSGVEPTITNSLDRSTENKNGRTANYIFVRAHDSVQTVIADI IREININPTDGLTF 653
QY 587 TWEETKKAFAEYNKDLLATEKKYTHYNTALSYALLLTNKSVPVRYVYDGMFDDGQYMAH 646
Db 654 TMDELKQAFKIYNEEDMRKADKKYQFNPTAHALMSLNKDSITRYVYGDLYTDDGQYMEK 713
QY 647 KTINVEAETLTKARIKTVSGGQAMRNQOVG-----NSEIITSVRYGKALKATDT 697
Db 714 KSPVHDAIDALLRARIKTVAGQDMKVYMGVPREADKWSYNGILTSVRYGTGANEATDE 773
QY 698 GDRTRTSGVAIEGNPNSRLKASDRVVVNVNGAAHKNQVREPLLTITDNGIKAVHSDOE 757
Db 774 GTAETRTQGMVIAASNPNLKLNEWDKLQVNVNGAAHKNQYRPLVLTITKDGISRLYTDSE 833
QY 758 A-AGLVRYTNRGELIFTAADIKGYANPQVSGYGVVYVPGVGAALIKMFALRLRPHQWA 816
Db 834 VPQSLWKKITDANGILTFDWNIDAGVSNVQVSGYLAIVVPGAKENQDARVATSKKXNASG 893
QY 817 SVHQNA-ALDSRVMEFGESNFQAFATKKEEYTNVVIKAVNDKFAEAGVTDFFENAPQYVSS 875
Db 894 QVYESSPALDSQLIYEGESNFQDPATRODQYTNKVIKAVNLFKEWGVTSFELPPQYVSS 953

```
QY 876 TDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQWY 935
Db 954 QDGTFLDSIIQNGYAFEDRYDAMSNNKYGSLDNLALRALHSVNTQAIADWVPDQIY 1013
QY 936 AFPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAFLELOAKYPBL 995
Db 1014 NLPGEVVTATRVNNGYTYREGAEIKEKLYVANSKTNGTDYCKYGGAFLELAKYPEI 1073
QY 996 FARKQISTGVPMDPSVKIKQMSAKYFNGTNIILGRGAGYVLKQDAQNTYFNISDNKEINFL 1055
Db 1074 FERVQISNGQKMTDEKIKWSAKYFNGTNIILGRGAYVLKDWGSKY-LSNKNGETAL 1131
QY 1056 PKTLNQDSQVGSFDGKGYVYVYSTGSQAQKNTFI-SEGDKWYYPDNNGYVMTGAQSING 1114
Db 1132 PKQLYNKEASTGFVXDTNGFKFYSTSGNQAKDTFIQDENGWNYYPDNQGYLVTGAREIDG 1191
QY 1115 VNYVFLSNGQLRDAILKNEIDGTAYYGNDRRYENGYYQFMGQVWRHFN-NGEMSVGLT 1173
Db 1192 KQLYFMKGVQURDALQEDENGQYVYDKTGAKVLNRYYTSDQNRWYFDAGKVMARGLV 1251
QY 1174 VI-DGQVQYFDMGVOAKGKFTTTADGKIRYDFDKOSGNMYRNFTEENEE-GKWLYLGBDG 1231
Db 1252 KIGDQ-QYFDQNGYQVKGKVRRAKDGKLYDFDKDSGNVAVINRFAQGNPSPDWYVFGADG 1310
QY 1232 AAVTGSQITNGHLYFRANGVQVKGFEVTDHGRISYVDGNSGDIQNRFRVNAQGWYF 1291
Db 1311 VALTGKIQGQTLYFGDGKQVQVVLADKSIYFDANSSEMAMVKNKFAEGAKNEWY 1370
QY 1292 FDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDRIYGRISYVDGNSGDIQNRFRVNAQ 1351
Db 1371 FDDGKAVTGLKTIINQVLYFDQDGKQVQVVLADKSIYFDANSSEMAMVKNKFAEGAK 1430
QY 1352 GQWYFDNNGYAVTGARTINGHLYFRANGVQVKGFEVTDHGRISYVDGNSGDIQNRFR 1411
Db 1431 NEWYFDQDGKAVTGLQTKINKVLYFGDGKQVQVVLADKSIYFDANSSEMAMVKNK 1490
QY 1412 VRNAQGWYFDNNGYAVTGARTINGHLYFRANGVQVKGFEVTDRIYGRISYVDGNSG 1469
Db 1491 AEGAKNEWYFDQDGKAVTGLKTIINQVLYFGDGKQVQVVLADKSIYFDANSSEMAM 1547
QY 1470 ERVR 1473
Db 1548 DMVR 1551

RESULT 9
Q69A94 LEUME
ID Q69A94_LEUME PRELIMINARY; PRT; 1454 AA.
AC Q69A94;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Dextranucrase (EC 2.4.1.5).
GN Name=drrp;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IBT-PQ;
RA Fernandez-Piqueres J.L., Lopez-Munguia A., Olivera C.;
RT "Molecular characterization of a dextranucrase gene from Leuconostoc
RT mesenteroides IBT-PQ isolated from pulque."
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY504865.1; AAS79426.1; -; Genomic DNA.
DR GO: 0047849; F:dextranucrase activity; IEA.
DR GO: 0046757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO: 0009250; F:glucan biosynthesis; IEA.
DR InterPro: IPR002479; CW binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; CW binding_1; 1.
DR Pfam: PF02324; Glyco_hydro_70; 1.
```

```
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;
Query Match 47.2%; Score 3650; DB 2; Length 1454;
Best Local Similarity 52.0%; Pred. NO. 1.6e-161;
Matches 755; Conservative 210; Mismatches 398; Indels 90; Gaps 31;
QY 17 VTVSVASAVMTLTLSGLVKADSNESKSOISNDSNTSVVTANEESNVITATSQEA-- 74
Db 32 VTSFIQQAQADVQKNGVVTAVNQSNDSATTTDKS--ITTNDKATTTADTSTNDKATT 89
QY 75 -ASSQTTHVTTSSTSVVNPKEVSNPYTVGETASNGEKLQNTTVD-----KT 125
Db 90 TADTSTNDKATTTADTST--NDKAT-----TTADISTNNK-----ATTTADTSTNNKAAT 138
QY 126 SEAAANNISKOTTEAD-----TDVIDDSN-----AANLQILEKLPNVKEIDGKYVYD 173
Db 139 ADTSNNNSATTSDDKVSLSLAKSQSIDNNSKTDTDTAASFEASSK--NLKTIIDGKYVYD 196
QY 174 NNGKVRNFTLIADGKILHFD-ETGAYTDTSIDTVNKDIVTTRSNLYKYNQVYDRSAQS 232
Db 197 DNGQIKGNFATVIDGKLYFDKDTGALADTN-DYQFLEGLTSENNNYTEHNASVGTSSAS 255
QY 233 FEHVDHYLTABSWRYPKYLKDGKTWTOSTEKDPRPLMTWMPDQETQRYVYVYNAQLG 292
Db 256 YTNVDGYLTADSWRYPKDI FVNGQNWESSKDDDLPLMTWMPDKAT---HVNYLNAMY 312
QY 293 INKT-----YDDTNSQLQNLIAAATIQAIEAKITTLKNTDMLRQTIISAFVKTSQANS 347
Db 313 LDATETVTVTSDSDQDALNKAQNIQVKEIKESQEVOTQWLKDDISKFVDSQSNWNIA 372
QY 348 SEKFPDDLHONGAVLYDNEGKLTYPANSNYRIILARTPTNQTGKDPRTADNTIGGYEPL 407
Db 373 SESKGTDLHOGALLYVNSDK--TPDANSDYRLARTPTNQTG--TPLYTPTPTQGGYDPL 429
QY 408 LANDVDSNPVVQAEQLNLHLFLMNFNIYANDPDANFDSIRVDADVNDVADNLDLQIAGDY 467
Db 430 LANDVDSNPVVQAEQLNGMYLLNFSGITNNDADANFDSIRVDADVNDVADNLDLQIADY 489
QY 468 LKAARGHKNDKAANDHLSTLEAWSNDNTPYLHDDGDNMNMNNDKRLSLFLSLAKPL-- 525
Db 490 FKAAYGVGDKSDAISNQHVSIILEDSDNDAEYVKGNDQNLSDNKLRLSLKSLYSLTPPVD 549
QY 526 ---NORSGMPLITNSLVNRTDDNAETHAAPSYSFIRAHDSVEVDLIADIKAENPNVV 582
Db 550 HYGKRSGLPEFLTNSLVNRTNDSRDNTAQNYSFVRAHDSVEQTVIAEIIKQRIIDPSD 609
QY 583 GVSFTMBERI KKAFLYINKDLALATEKCYTHYNTALSYALLTNKSSVPRVYVYGMFTDDQ 642
Db 610 GLSPTMDQLTEAFKLYNADQLKTDEFTQYNI PSTYATILTNKDTVPRVYVYGMFTDDQ 669
QY 643 YMAHKNTINYEAIETLLKARIKYVSGGQAMRNQV--GNSE-----TITSVRYGKGALK 693
Db 670 YMATKSLYDAIDTLTKSRIKYVSGGTMSMKYMGQSDSSMAADSRYGILTSVRYGNGAMT 729
QY 694 ATDTGDRTRTSGVAVIEGNNPSRLKASDRVVVNMGAHKNQAYRPLLLTTDNGIKAYH 753
Db 730 ATDAGTNETRTHGIAVIESNNPDLKLSSTDQVVDVMDGIAHKNQAYRPLALLTKDGIDTV 789
QY 754 SDQEA--GLVRYTNDRGELIFTAADIKGYANPQVSGYLVGVVVPVGAALI KMPALRLARPH 812
Db 790 SDSVDSGLIRYTNNGQLIFNSSDITVGTAAPQASGLAVAVPVGAS-----DTQARTE 844
QY 813 QQMA-----SVHQNAALDSRMFEGFSNFQAFATKKEEYTNVVIKAVNDKFAEAGVTDF 866
Db 845 SSTATTTDGTQLHSNAALDSQVIVESFNSFQSTPTTEAEYANVQIANNTDLYKSWGITNF 904
QY 867 EMAYQVVSSTGCSFLDSVIQNGYAFTRDYDLIGISKPNKYGTADDLVKAIKALHSKGIKVM 926
Db 905 EFPQYRSSTDSFLDSIIQNGYAFTRDYDLIGFNTPTKYGTVDQLRTAKALHATGIKAM 964
QY 927 ADWVPDQWYAFPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAFLE 986
```

Db 965 ADWVPQIYNLTGKEVAVQVRVNNSGIYNQDSVINTKLTVASQTVGGG-EYQALYGGFELD 1023
QY 987 ELQAKYPELPARKQISTGVPMDFSVKI KQWSAKYFNGTINILGRGAGYVLKQDQATNTYFNI 1046
Db 1024 EIKKLYPSLFEXKQISTGVPMDESEKI KWSAKYFNGTINILGRGAGYVLKQDQATNTYFNI 1083
QY 1047 SDNKEIN-PLPKTLNQDSQVGF-SYDGGYVYYSYGQAKNTFISEGDK--WYFDDNN 1102
Db 1084 STSSNSVFLPKQLTNEESNTGFISTDG-GMTYISTSGYQAKDTFTQD-DKSNWYFDDKN 1141
QY 1103 GYVMTGAQINGVWYFLSNGLQRLDAILKNEDGTVAIYNGDGRYVYQFMSGVWRH 1162
Db 1142 GYMTYGFQVNDNNYFLPENGIELQDAILEDKGDVYVYFQYKQITIDGYMLANKTWRY 1201
QY 1163 FN-NGEM-SVGJT--VIDQO--VQYDEMGOAKGFVTTADGKIRYFDKQSGNMYRNF 1216
Db 1202 FDKNGVMANAGLTVTVDGQKHQYFDKNGIQVKGTSVKDADGKLYFDTDSGEMVYRNF 1261
QY 1217 IENESGKWLILGEDGNAVTSQTINGOHLIFRANGVOVKGFEVTDHGHRIISYDGNQSGDQ 1276
Db 1262 GSKTGTWTSYFGADGIAVTGAQITISQKLFDFADGQIQKGEATDKKGHGHYTDADSGEM 1321
QY 1277 IERNFRVNAQOWFYFDNNGYAVTGARTINGOLLIFRANGVOVKGFEVTDHGHRIISYDGN 1336
Db 1322 TTRREKLSGDSWYFNKGNVVTGAQVINGOHLFFESNGNQVKGREYATDQKMYIDA 1381
QY 1337 NSGDIQIRNRFRVNAQOWFYFDNNGYAVTGARTINGOHLIFRANGVOVKGFEVTDHGHRI 1396
Db 1382 DSGDMVYRFRISDGSWAYFGANGVAVTGAQINGOHLFFESNGNQVKGREYATDQKMYIDA 1441
QY 1397 SYDGNQSGDQIRN 1409
Db 1442 KYDANSGBELIKS 1454

RESULT 10
Q9LCH3_STOR PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfR;
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1303;
RW [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC10557;
RX MEDLINE=20231779; PubMed=10768934;
RX DOI=10.1128/TAI.68.5.2475-2483.2000;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RL encoding glucosyltransferase from Streptococcus oralis.";
RL Infect. Immun. 68:2475-2483(2000).
DR EMBL; AB025228; BAA95201.1; -; Genomic_DNA.
DR HSSP; P06853; 1HCX.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; 1CW binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26B4D7C2E543 CRC64;
Query Match 46.5%; Score 3597.5; DB 2; Length 1575;
Best Local Similarity 48.0%; Pred. No. Se-159;
Matches 749; Conservative 236; Mismatches 475; Indels 101; Gaps 24;
QY 1 MDKVRVYKLRVKKRWVTVSVASAVMTLTT-----LSGGLVKAD----- 39

Db 2 MEKIHVKHVKQKQVVAIALATLALIVAPKVLGLSEGVVHADDDVQVNVQEPATQTS 61
QY 40 ---SNEKSOISNDNTSVVTANEESNVLTATEATSKQEAASSQTNHVTSSSS--TSVNP 95
Db 62 PQOQTPAQAKIASQEAQEKVTPADKVTDDVAASEKPAKPAENTEATVQNAQPAKPAD 121
QY 96 KEVSNPYTVGTASNGEKLQQTITVDKTSBAAANNISKQTTTEADTD-----VIDDS 148
Db 122 KEASTEKAAVAEVEV---KAANAITEIPKTEVADQNKQARPTTAQOQEGDKREKTAVE 177
QY 149 NAANIQLILEK-----LPNKVIDGKYYVD 173
Db 178 IVANPKVAKKDLPLBPFGSKQGAIAERWADQAPAPVADHDDDLVLSHKTIDGKNYVQ 237
QY 174 NNGKYRTNFTLLADGKILHFD--ETGAYTDTSDTVNKKDITVTRSNLYKKYNOVYDRSAQS 232
Db 238 DDTGTVKQNFVAVELNGKILYFADTETGALVDSN-EYQFQQTSSLNNEFSQKNAPYGTDDK 296
QY 233 FEHVDHYLTAEWYRPKYLKDGKTWTOSTETKDFRPLMTWMPDQETQRYVYNNMAQ-L 291
Db 297 IETVDGYLTADSWYRPKFLKDGKTWTASTETDLRPLLMWMPDKRTQINLYNNMQOGL 356
QY 292 GINKTYDDTSNLQNLIAAATTOAKIEAKITTLKNTDMLRQTIISAFVKTOSAWN----SD 347
Db 357 GAG-AFENKVEQALLTGASQQVQRKIEEKIGKEGDTKMLRTLMGAFVKTQPNWNKTES 415
QY 348 SEKPFDDHLQNGAVLYDNEGKLTVPANSNYRILNRTPTNCTKPKOPRYTADNTIGYFEL 407
Db 416 TGTGKDHQGLLYTNEK-SPHADSKFRLLNRTPTSQTG--TPKYFIDKSNNGGYEFL 472
QY 408 LANDVDNSPVVQARQLNWLHFLMNFNIYANDPDANFDSIRVDVNDVNDADLLQAGDY 467
Db 473 LANDFDNSPVAQAEQLNWLHMYMNFSGISIVANDPTANFDGVVADVADVNDVNDLQIASDY 532
QY 468 LKAAGIHKNDKAANDHLSILEAMSDNTPYLLHDDGDNMNMNKLRLSLPLSLAKPLNQ 527
Db 533 PKSRVYKVGESBEEAIKHLISILEAMSDNDPDYKDKTGAQLAIDNKLRLSLYSFMRNLGI 592
QY 528 RSGMPLITNSIVNFTDDNAETAAPVPSYFTRAHDSQVODLADIKAENINNVGYST 587
Db 593 RSGVEFTITSLNDRSSEKNGERMANYIFVRAHDSQVQTVIADIIRINENPTDGLTFT 652
QY 588 MEEIKKAFETYNKOLLATEKTYHTNTALSYALLTNKSSVPRVYVYGMFTDDGQYMAHK 647
Db 653 MDLQKQAFKIYNEDMRKADKKYQFNIPTAHALMLSNKDSITRVYVYGLYTDGQYMEKK 712
QY 648 TINYEAITLLKARIKYVSGGQAMRNQVQV-----NSBITSVRYGKALKATDTG 698
Db 713 SPYHDAIDALLRARIKYVAGGQDMKVTYMGVPREADKWSYNGILTSVRYGTGANEATDSG 772
QY 699 DRTTTSQVAVTEGNNPSRLKASDRVVVNGGAHKNQAYRPLLLTTDNGIKAYHSDQBA 758
Db 773 TAETRTQGMVAVIASNNPMLKLNWDKLCQVNNGAHKNQYRPLVLTTKDGI SRYLTDEEV 832
QY 759 -AGLVRYTNDRGELIFTAADIKGVANPOVSGVLGVVVPVGAALIKWFAALRLARPHQOMAS 817
Db 833 PQSLWKTGDANGILTFDMDIAGVSNQVSVGLAVVVPVGAQADQARTTASKKQNASQ 892
QY 818 VHQ-NAALDSRVMPFEGFSNFOAFATKKEEYTNVVIKNDKPAEWGVTDFFEMAPQVYST 876
Db 893 VYESSAALDSQLIYEGFSNFQDFATRDDQYTNKVIKAVNVLFKEWGVTSEFELPPQYVSSQ 952
QY 877 DGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWDVQMYA 936
Db 953 DGTFLDSIIQNGYAFEDRYDMANSKNNKYGSLKOLLNLRALHSVNIQAIADWVDPQIYN 1012
QY 937 FPEKEVWTATRVKDYGTPVAGSOIKVTLVYVDGSSGDKQQAQYGGAFLEELQAKYPELF 996
Db 1013 LPGEVWTATRVNNYGYTYREGAEIKELVYVANSKTNETDFQKYGGAFLEELQAKYPELF 1072
QY 997 ARKOISTGVPMDFPSVKIKQWSAKYFNGTINILGRGAGYVLKQDQATNTYFNISDNKEINFLP 1056

```
Db 1073 ERVQISNGKMTTDEKITKASAKYFNGTNILGRGAYVYLKDWASNDYLT-NRNGEI-VLP 1130
Qy 1057 KTLNQNDSQVFSYDGKGVYVYSTSGYQAKNTFI-SEGDKWYFNNNMYMTGAOSINGV 1115
Db 1131 KQLVNKNSYTGFSVSDANGTKFYSTSGYQAKNSFIQDENGWYFFDKRGYLVGTGAHEIDGK 1190
Qy 1116 NYFELSLGLORDALTKNEDGTAYYNGDGRYENGYYQPMGCVWRHEN-NGEMSGLTV 1174
Db 1191 HVFUJNGIQLRDSIREDENGQYYYDTGQAVLNRYYTTDQGNWRYFDAGVWARGLVK 1250
Qy 1175 I-DGQVQFDEMGYQAKGFVTTADKIRYFDKQSGNNRNF1ENE8-GKWLVLGEDGA 1232
Db 1251 IGDGQ-QFDENGYQVKGKIVSAKDGKLRYPDKDSGNVINRFPAGDNPDMWYFVGEVFA 1309
Qy 1233 AVTSGQTTNGHLYFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFVNAOQWYF 1292
Db 1310 KLTGLQKIGQOTLYFDQDGKQVKGKIVLSDKSIYFPDANGEMA VGFKAEGAKNEWYF 1369
Qy 1293 DNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDGNSGDQIRNRFVNAOQ 1352
Db 1370 DKTGAVTGLQKIGQOTLYFDQDGKQVKGKIVLSDKSIYFPDANGEMA VGFKAEGAKN 1429
Qy 1353 QWYFDNNGYAVTGARTINGHLYFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFV 1412
Db 1430 EWFYFDQTKAVTGLQKIDQTLYFDQDGKQVKGKIVLSDKSIYFPDANGEMA VTKFV 1489
Qy 1413 RNAQSQWYFDNNGYAVTGARTINGHLYFRANGVQVKGFEVTDYGRISYYDANGSERV 1472
Db 1490 EGSQNEWYFDQAGKAVTGLQVQGTLYFTQDGKQVKGK-VVDVNGVSYRFDANGSDWA 1548
Qy 1473 R 1473
Db 1549 R 1549

RESULT 11
Q9WXJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GTP-S.
GN Name=gtpf;
OS Streptococcus criceti.
OC Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1333;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSSP; P06654; IMPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1_2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 45.5%; Score 3523.5; DB 2; Length 1512;
Best Local Similarity 48.0%; Pred. No. 1.3e-155;
Matches 739; Conservative 239; Mismatches 459; Indels 101; Gaps 34;

Qy 1 MDKRYKLRKVKRWVTSVASVMTLTLSGG---LVKADSNESKSQISNDNTS--- 54
Db 1 MERKLRKLVKVKQWVTIAVASGLASVIGAGASQTVSADDIGNGASASAEQNTSASQ 60
Qy 55 ---VVTANEESNVTIETASKQEAASSQTNHTVTTSSSTSVSNPKVSVN-----PYTVG 106
```

```
Db 61 NKEVDSDAQAQDAKATSEQAIVSS-----VDTNSEIDQVNDVGVSAANNQPEAPAPQ 115
Qy 107 ETASNG-----EKLQNTTTVDKTSAAANNISKQTTEADTDVDDSSNAANLQILEKLPN 161
Db 116 AAASNNNTATSEBANTNTAVSEAAPAAEN-----RTAEKGADLSQDEAAAL-----SLDN 167
Qy 162 VKEIDGKYYYDNGKVRTNFTLADGKILHFD-ETGAYTDTSDITVKNKDIVTTFSNLYK 220
Db 168 IKKYDGYKYVMADSGYKKNFAITVKGQLLYPDAETGALSTSTSYFSQGLTPLVSD-FS 226
Qy 221 KYNVYDRSAQSFHVHLYTAESWYRPKYILKQKWTQTQTEKDFRPLLLMTWPDQETQ 280
Db 227 INNKAFOSSAKSFELVDGYLTAESWYRPTKLENGKTWDSKETDLRPLVTSWNPDKDTQ 286
Qy 281 ROYVNYNMAQIGINKTYDDTSNQLNIAAATIQAIEAKITTLTKNTDMLRQTISAFVKT 340
Db 287 VAYLNTYSKALGGKEEFTTKTSQATALNTAAEWIQWIEORISKEQGTAWLRDAMAFAVAT 346
Qy 341 QSAWNSSEKPPD--DHLQNGAVLYDNEGKLTTPYANSYRILNRTPTNQTQKQDPRYTAD 398
Db 347 QSRWNSDSEQ-FDKNDHLQGGALLYTN-NKLTWADSKYRLNRTPTRDGK--PHYSKA 402
Qy 399 NTIGSYEFLLANDVNSNPVQASQLNWLHFLMNFNIYANDPDANFDSIRVDAVDNDA 458
Db 403 DEYGYEFLLANDVNSNPVQAEMLNQIHLYMNGSLVMGDKNANFGIRVDAVDNDA 462
Qy 459 DLLQTAGDYLKAAKGIHKNKKAANDHLSILBEASDNDTPYLHDDGDNMNMNKLRLSLL 518
Db 463 DTQLQYTNYPNAVYGVDSKEAQAHAHISILEAWSYNDNYNQDTNGAALAMDGLRLSLL 522
Qy 519 FSLAKPLNQRS-GMNPLITN--SLVNRDADNAETAAPVPSYFIRAHDSVQDLADIKA 575
Db 523 YTLRPLSERTPGLSTLTKSEYGLTDRTKDKYGDTPSYVYFVRAHDSVQTVIAQIIKE 582
Qy 576 EINPVVGYSFTEBEEKKAFIYNKDLATATKYYHTNTALSYALLLNKSSVPVRYGD 635
Db 583 KIDPTDGTFTLDQLKQAFDIYNKDMNSVBKHYTHYNI PAAYAVMLNMSVTRVRYGD 642
Qy 636 MFTDDGQYMAHKTINYEAETLLKARIKYVSGGQAMRNQO-----VGN 678
Db 643 LFTDDGQYMETKSPYDAINTLLRARIRYAGGQTMHKA YTPSAAMKAKNPDSGVLGN 702
Qy 679 SEITSVRYGKALKATD-TGDRTRTSGVAVIEGNPSLRKASDRVVVVMGAAHKQQA 737
Db 703 SEVLVSVRFGQVMSADDMTGGLAKTSGMFLSIANNPELELDANEETKVNVGKIHAGQA 762
Qy 738 YRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVWVPVG 797
Db 763 YRPLLLTTDGLQKLYLSDS-TNLTKVADKQGFITFKGSEIKGYQVGVNGVLSVWVPVG 821
Qy 798 AALIKMFALRLARPHQQMASVHONA-----ALDSRVMEFGFSNFQAFATKBEY 847
Db 822 AKSDQ--DIRVA-----ASTKANAKGDKSVTASQALDSQLIYEGFSNFQDFVQKDAQYT 873
Qy 848 NVVIAKNVDKFAEWGVTDFEMAPQVSVSDSGFLDSVLQNGYAFATDRDLGLSKNKYGT 907
Db 874 NKKTAENTDLFKAWGVTSFEMAPQVSVSATDGTFLDSIIQNGYAFSDRYDLAMSKNKYGS 933
Qy 908 ADDLVKALKALHSGIKVMADWVPDQMYAFPEKEVVTATRVDKYCTPVAGSIOKNTLYV 967
Db 934 KEDLANALKALHAAGIQAIADWVPDQIYQLPGKEVVTASRVNDRVGRVKIDQPMVWKLIA 993
Qy 968 DGKSGKQQQAKYGGAFLEELQAKYPELFARQKISTGVPMDFPSVKIKQWSAKYFNGTNIL 1027
Db 994 NTKSGKDFQAKYGGEFLEELQKYPEMFTAKMISTGKPIDSSVLKEMSAQYFNGTNVL 1053
Qy 1028 GRGAGYVLKQATNTYFNI SDNKEINFLPKTILL-NQDSQVGSYDGKGVYVYSTSGYQAK 1086
Db 1054 GRGTDYVLSDSGTGKYFTVNEKGE--FLPAVLTDGDEAKTGFYNDGKGMGYFTTAGSQAK 1111
Qy 1087 NTFISEGDKWYFDNNGYVMTGAOSINGVNYFFLSNGLQLRDAILKNEIDGTVAAYVNDGR 1146
```



```
Db 1112 SDFVTVAGNTYFDYTGHMVMTGNGINTKPYFVFLPNGVMLKDAVMBDDRGSRVYVYKGTG 1171
Qy 1147 RYE---NGYOFMSGV---WRHFN--GEMSVGLTVIDGOVQVQFDEMGYQAKGKEVTTA 1197
Db 1172 MYKSRNNEFWAMTDSKGQLRFRHFDNYGFMSVGLVTIGNVQYDEEGFQVKGDFVTK 1231
Qy 1198 DKIRYFDKQSGNMVNRRIENEEGKWLJGBDGAATVGSQTINGQLHYFRA-NGVQVKG 1256
Db 1232 AGCTRYFDKNTGNLVKGOFF-NONGHWYSDOGLIAKGAQTIKGOKLYFDKATGAQVKG 1290
Qy 1257 EFTVHHGRISYDGNSGDQIRNRFVRNAQOGWFYFDNNGYAVTGARTINGQLHYF-RAN 1315
Db 1291 DFVTDKGNFTFYSGTGDGLAVSTFPSTGNNAWFYADENGVAKGKTINGOKLYPDTKT 1350
Qy 1316 GYVQKGEFTDGRISYDGNSGDQIRNRFVRNAQOG--QWFFVFNNGYAVTGARTINGQ 1373
Db 1351 GQOAKGRFVRDAKG-LRFYDAUTGALVTNSFLETRAGSNQWYMGADGYAVRGHQTIOQR 1409
Qy 1374 HLYFRA-NGVQVKGFBVTDHGRISYDGNSGDQIRNRFVRNAQOGWFYFDNNGYAVTGA 1432
Db 1410 HMYFDAETQQAQGIWVTDANGRKVFYDANTGDRVNVQFVL-VNGSWYFFGYDGAATGF 1468
Qy 1433 RTINGOHLYFRANGVQVKGFEFTDGRISYDGNSSGE 1470
Db 1469 RDIRGOHLYFNPDPGTOAKGTTVKID-NRIYTFDADSGE 1505

RESULT 12
Q56CX8 95TRE PRELIMINARY; PRT; 1506 AA.
AC Q56CX8;
DC Q56CX8;
DT 10-MAY-2005 (TremBLrel. 30, Created)
DT 10-MAY-2005 (TremBLrel. 30, Last sequence update)
DE Glucosyltransferase-T.
GN Name=gftt;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR2005) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY966490; AAX76986.1; -; Genomic_DNA.
KW Transferase.
SQ SEQUENCE 1506 AA; 167016 MW; 94B882EF2C17C451 CRC64;

Query Match 45.4%; Score 3512.5; DB 2; Length 1506;
Best Local Similarity 46.5%; Pred. No. 4.3e-155;
Matches 716; Conservative 262; Mismatches 461; Indels 101; Gaps .31;

Qy 1 MDKVVRYKLRKVRWTVSVASAVMTLTTLSGGL-----VKADSN 42
Db 1 MERKLHYKLHKVKKQWTVIAVASAGLSTVAGSLSQTVSADDLAKEQAQAAASQQAANAQ 60
Qy 43 SKSQISND-SNTSVVTANEESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSN 101
Db 61 NEDEVASDAADTASAKATSEKEVWQSSDTNSQVETKQAKESADAVAK-QAPQAG 119
Qy 102 PVTGVTASNGEKLQNTTVDKTEAANNISKQTTADTDVIDDSNAANLQILEKLPN 161
Db 120 PATTSQVASSSSSVAPSEADK---AAGSVSQNEEAAAL-----SLAN 161
Qy 162 VKEIDGKYTYDNNNGKVRNFTLIADGKILHFD-ETGAYTDTSDIVNKNRDIWTRSNLYK 220
Db 162 IKKIDGKYTYVWADGSGYKKNFAITVDGQMLYFDKATGALSSTSTVSFSGGLTPIVSD-FS 220
Qy 221 KYNQVDRSAQSFEHVDHYLTAESYRKYILKQGTWQSTQTEKDFRPLMTWPDQSTQ 280
Db 221 VNNKAFDSSEKSFELVDGYLTAESYRPAKILENGKTVWDSKETDLRPLMSWNPNDQ 280
```

```
Qy 281 ROYVNMNAQLGINKTYDDTSNQLQINIAAATQAKIEAKITTLKNTDMLROTISAFVK 340
Db 281 VAYLVNMSKALGKKEEFTTETSQLTLNTAAELIAQAKIEARVSKEQCTKWLREAWAFVAT 340
Qy 341 QSAWNSDSEKPPD--DHLQNGAVLYDNBGLKLPYANSNRIILNRTPTNQTGKDKPPYTAD 398
Db 341 QSRWNKDSQ-YDKADHLOGGALLYTN--NNLTSEWANSWRLNRTPTRODGK--THYSKA 396
Qy 399 NTIGGVFEFLANDVNSNPVQAEOLNWLHFLMFGNIIVANDPDANFDSIRVDADVNDVA 458
Db 397 DKYGVGEFLANDVNSNPVQAEMLNQIHYLMNWEIVMGDKNANFDGIRVDADVNDVA 456
Qy 459 DLLQIAGDYLLKAAKGHKNDKAANDHLSILEAWSNDPTYLHDDGDNMINMDNKLRLSL 518
Db 457 DTLLQLYTNYFNSVYGVNKSQAALAHISVLEAWSYNDNDYNDQDTNGAALAMDNGLAFSL 516
Qy 519 FSLAKPLNQRS-GMPLITN--SLVNRDNDNAETAAPSPYSFIRAHDSVQDLIAIICA 575
Db 517 YTLTRPINERTPGMSTLIKSEYGLTDRTKNDKYGDTQPSYVFRVRAHDSVQTVIAQIIKE 576
Qy 576 EINPNVGVSYFTMBEIKKAFELYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVYGD 635
Db 577 KIDPTDGTFTFLDQLQKAFELYNKMNSVNKHYTHYNIPAAAYAVMLSNMESVTRVYGD 636
Qy 636 MFTDDGQYMAHKTINYEAITLLKARIKYVSGGQAMRNOQ-----VGN 678
Db 637 LFTDDGQYMAKSPYDAINTLLRARIYAAGQIMEHNSYKPSAAMKAHPDAGNVGN 696
Qy 679 SEIITSVRYGKALKATD-TGDRTRTTRTSGVAVIEGNPSLRKASDRVVVNNNGAAHNOA 737
Db 697 SEVLVSVRFGQDMSADDMTGKLAKTSGMFTLISNNPELELDVNEEIKVNVGKIHAQOA 756
Qy 738 VRPLLLTTDNGIKAVHSDQEAAGLVRYTYNDRCELFTADIKGYANPOVSGYLVGVVPG 797
Db 757 VRPLLLTTDKGLQKYLNDSDTK-LTKIADKQGITFKGSEIKGYQKQVNGVLSVWVPG 815
Qy 798 AALIK--MFALRLARPHQOMASVHQAALDSRMFEGFSNFQAFATKKEEYTNVVIKXV 855
Db 816 AKADQDIRVAPSTAAGEKAKTYTASQALESQLIYEGFSNFQDFVQKDSQYTNKKIAENT 875
Qy 856 DKFAEWGVTDFEMAPQYVSSDTSGLDSVQNGYAFTRDYDIGISKPNKYGTADDLVKAI 915
Db 876 DLFRANGVTSFEMAPQYVSATDGTGLDSIENGYAFTRDYDLAMSKNNKYGSKEDLANAL 935
Qy 916 KALHSGIKVMADWVPOQWYAPPEKEVVTATRVDKYGPVAGSQIKNTLYVVDGSSGKD 975
Db 936 KALHAAGIQAIADWVPOQIYQLPGEVVTASRVNDYGRVKVDQPLVEKLYLANTKSSGKD 995
Qy 976 QOAKYGGAFLEELQAKYPPELFARKQISTGVPMDPSPVKIKOWSAKYNFNGTNILGRGAGYVL 1035
Db 996 FOAKYGGFLEELQAKYPEMFTTKMISTGKTIDPSVKLEWSAKYFNGTNVLDRTDYIL 1055
Qy 1036 KDAQNTVFNISDNKBINFLPKTL--NODSQVGSYDGKGYVYVYSTGSAQKNTFISEGD 1094
Db 1056 SDEGTGKYFTV--NEKGDFLPASLTGNKDAKTGFYNDGKGIYVYTTAGNKARSAFVTEAG 1113
Qy 1095 KWYTFDNNYVMTGAQSIINGVNVYFSLNGLQLRDAILKNEDGTAYAYGNDGRYENGYQ 1154
Db 1114 NTYYPDYTGHVMTGPNVINTKFFYFLPNGIMLKDAIKQDEKRSYVYKGTGYMYRGK---- 1170
Qy 1155 FMSGVW-----RHFNN--GEMSVGLTVIDGOVQFDEMGYQAKGKEVTTADGKI 1201
Db 1171 -RDNEFWAMTDSKGMGRFRHFDYGFMSIGLVTINQNVQYDENGFOVKGEFVTDQDGT 1229
Qy 1202 RYFDKQSGMYNRRIENEEGKWLJGBDGAATVGSQTINGQLHYF-RANGVQVKGFEFT 1260
Db 1230 RYFDQSGNVLKGOFL-NKDGWNYLDDQGLVAKQAQTIKGOKLYFDTKTGQVQKGFVT 1288
Qy 1261 DHGRI SYDGNSGDQIRNRFVRNAQOGWFYFDNNGYAVTGARTINGQLHYF-RANGVQV 1319
Db 1289 KDQGNTPFYSGTGDGLILGQFFSTGNNAWFADENGHVAKGAKTIRGOKLYFDTKTGQOA 1348
Qy 1320 KGEFVTDYGRISYDGNSGDQIRNRFVRNAQOG--QWFFVFNNGYAVTGARTINGQLHYF 1377
```



```
Db 1362 GROVKGHFVTTNNQRY-FLDGDGSEIAPSRFV-TENNKNVYVDGNGKLVKGAQVINGNHV 1419
Qy 1311 YFRANGVQVKGFEVTDYRGRISYDNGSGDQIRNFRVNAQOGWFYFDNNGYAVTGARTI 1370
Db 1420 YFNNDYSQVKGAWNGR-----YDGDGSGQAVSNQFIQIAANQWYALNQDGHKVTGLQNI 1474
Qy 1371 NGQHLVFRANGVQVKGFEVTDYRGRISYDNGSGDQIRNFRVNAQOGWFYFDNNGYAVT 1430
Db 1475 NNKVVYFGSNGAQVKGKLLTVQ-GKICYFDAHTGEQVNRVFEAARGCWYFNSAGQAVT 1533
Qy 1431 GARTINGOHLVFRANGVQVKGFEVTDYR--GRISYVDANSGE 1470
Db 1534 GQVINGKQLYFDGSGRQVKGRIY---YVGKRLFCDAKTGE 1572

RESULT 14
Q8KZL5_9STRE
ID Q8KZL5_9STRE PRELIMINARY; PRT; 1554 AA.
AC Q8KZL5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfU;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
ON NCBI_TaxID=1310;
RX MEDLINE=21958684; PubMed=11960691; DOI=10.1016/S0304-4165(01)00240-9;
RA Hanada N., Fukushima K., Nomura Y., Senpuku H., Hayakawa M.,
RA Mukasa H., Shiroza T., Abiko Y.;
RT "Cloning and nucleotide sequence analysis of the Streptococcus
RT sobrinus gtfU gene that produces a highly branched water-soluble
RT Glucan.";
RL Biochim. Biophys. Acta 1570:75-79(2002).
DR EMBL; AB089438; BAC07265.1; -; Genomic_DNA.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding 1; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ
SEQUENCE 1554 AA; 171676 MW; 6981BCC1DAE24A73 CRC64;

Query Match 44.4%; Score 3434.5; DB 2; Length 1554;
Best Local Similarity 46.6%; Pred. No. 1.9e-151;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;

Qy 1 MDKVRKYLKRVKRWVTVSVASAVMTLTSLGGLVKAD---SNESKQ----- 46
Db 1 MEKGLHYLKHVKYKCHWVTTAVAS-IGLVSLVGAGTVSAEDRKVANDTTAQTGVDTGQDQ 59
Qy 47 -ISDNTSVV--TANESNVITEATSKQEAASQTNHTVTSSSTSVNPKVWNSP 102
Db 60 ATTNDANTTTDTADQSAN-----TNQDQAGSDQSNQDQAKQDTAN----- 103
Qy 103 YTVGETASNGEKLOQT-----TVDKTS-----EAAANNISKQTTEADTDVDDNSAAML 153
Db 104 -TDRNQADNSQTDNNQATQDATSPATDGTSVQRDAANVATAADQEGQTAPSEQKSAAL 162
Qy 154 QLEKLPVKEIDGKYIYDNGKVRNTTLADGKILHFD-ETGAYTDTSDTVNKOIV 212
Db 163 ----SLDNNVKLIDGKYIYVQADGSKYKCNPAITVNGQMLYFSDTGALSSTSYFSQGT 217
Qy 213 TTRSNNLYKYNQVDRSAGSFEHVDHLYTABSWSYRPKYLKDGKTWTQSTGTEKDFRPLMT 272
Db 218 TNLVDDFSHNKAYDSTAKSFELNGLYFANSYRPAFLRNGQTWSEASNENDLRPVLMS 277
Qy 273 WMPDQETQRYVYNNVAQLGINKT-YDDTSNQLNLNIAAATTAQKIEAKITTLKNTDMLR 331
```

```
Db 278 WMPDQETQRYVYNNVAQLGINKT-YDDTSNQLNLNIAAATTAQKIEAKITTLKNTDMLR 337
Qy 332 QTISAFVKTQSANWSSEKFPD--DHLQNGAVLYDNEGKLTTPYANSNYRLNRTPTTNOTG 389
Db 338 TAMEAFVAAQPKNNMSTEN-FNKGDHLOGGALLYTN-SDLTPWANSDYRLNRTPTTQDQ 395
Qy 390 KKDPRYADNTIGGYEFLLANDVNSNPVQAEQLNWLHFLMFGNIYANDPDANPDSIR 449
Db 396 TK--KYFTSGEGGYEFLLSNDVNSNPVQAEQLNQLHLYMNGDIWMGDKANPDGVR 453
Qy 450 VDAVDNVADALLQIAGDYLLKAAKGIIHKNDKAANDHLSILEAMSDNTPTLHDDGDNNIM 509
Db 454 VDAVDNVNADLLQVYSNYFKDNYKVTDSANALAHISILEAWSLNDNQYEDNNGTALS 513
Qy 510 DNKLRSLFLSLAKPLNQRSGMNPILTNSL-VNRTDDNAETAAPVPSYFTRAHDSVQDL 568
Db 514 DNSRLTSLAVITKOPGQIDLSNLISESVNKRANDTAYGDTIPTYSFVRAHDSVQTV 573
Qy 569 IADIKAIEINPNVVGYSFTMEBIKKAPEIYNKDLATEKKYTHYNTALSVALLLTNKSSV 628
Db 574 IAKIVKEIDTNSDGYTFDLDQKDAFKIYNEDMAKVNKYTHYNTPAAYALLLSNMSV 633
Qy 629 PRVYGDMPDQGYMAHKTINYEAETLLKARIKYVSGQAMRNQOV-GNSEIITSVRY 687
Db 634 PRVYTGDLTYDDQGYMAKSPYDAIATMLOGRIAYVSGGSEBHVHKVNGNQLSSVRY 693
Qy 688 GKALKATDT-GDRTTTSQVAVIEGNPSLRLKASDRVVNMGAAHKKQARPLLLTDD 746
Db 694 GQDLMSADDTQGTDLTSRTSLGLVTLVSNDPNLDEL-GGDSLTVMNGRAHANQARPLILGT 752
Qy 747 NGIKAYHSDQEAAGLVRYTNDRGEIIFTAADIKGVANPOVSGVLGVVVPVGAALIKMFAL 806
Db 753 DGVQSYLKDS-D-TNIVKYTDANGNLTFTADDIKGSTVDMSGYLAWVVPVGA-K-DGQDV 809
Qy 807 RLARPHQQMA---SVHQNAALDSRVMEFGFSNFOAFATKKEBYTNVVIARNVDKFAEWG 863
Db 810 RVAADTNQKADGKSLKTSALDSQVIEGFSNFQDFANNADYTNKKAENADFFKGLGI 869
Qy 864 TDFEMAPQVVSSTDCSFLDSVLQNGYAFTRDRLDGISKPNKYGTADDLVKALKHSGKI 923
Db 870 TSFEMAPQVVSATDCSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANALKALHANGI 929
Qy 924 KVMADWVPDQMYAFPEKEVWTRVYDKGTPVAGSQIKNTLVVDGSKSGKQQAQYGA 983
Db 930 QAIADWVPDQIYQLPGEVWTRVYDKGTPVAGSQIKNTLVVDGSKSGKQQAQYGA 989
Qy 984 FLEELQAKYPFELFARKQISTGVPMDDPSVKIKOWSAKYFNGTNIILGRGAGYVLKQDQATNTY 1043
Db 990 FLDELKAKYPDMFTVMNISTGKPIDPSTKIKQWEAKYFNGTNIILGRGAGYVLKQDQATNTY 1049
Qy 1044 FNISDNKEINFLPKTLL-NQDSQVGSFVDGKYVYVYSTSGYQAKNTFTISEGDKWYFDNN 1102
Db 1050 FTWNENG-D-FLPASFTGDQNAKTGYDGTGMAYYSTSGNKAIVNSFIYEGGHYFYFDK 1107
Qy 1103 GYMTGTA-QSINGVNYVFLSGLQRLDALTKNEDGTIAYGNDGRYB-NGYQFM---- 1156
Db 1108 GHMVTGSAEDDNDYIFLPNGIQMRDAIYDQAGNSYTYGRTGILYKGNMYPFVDPNN 1167
Qy 1157 --SGYRHFN-NGEMSVGLTVIDGQVYFDENGQYQAKGKFVTTADGKIRYFQKSGNMRY 1213
Db 1168 ANKTVRYFDANNVVAIGVRNMYGQTYFYFDENGQYQAKGQLLTDKG-THYFDEDNGAMAK 1226
Qy 1214 NRPTEBEGKYLGEDGAATGSGTNGOHLIYFR-ANGVQVKGFEVTDHGRIRISYDGN 1272
Db 1227 NKVF-NVGDDWYMDGNGNAVKQGPVANNQILYFNPETGVQVKGQFITDAQRTSYDAN 1285
Qy 1273 SQDQIRNF-----VRNAQOGWFYD-NGYAVTARTINGQL 1309
Db 1286 SGALKSSGFFTPNGSDWTYAENGYVYKGFQVAENQDQYIYFDQTTQKQKAARVDGRD 1345
Qy 1310 LYFRA-NGVQVKGFEVTDYRGRISYDNGSGDQIRNFRVNAQOGWFYFDNNGYAV- 1364
```

```
Db 1346 LYFNPDSGVQVKGDFATDESNGTSTFVHGDNKGKVVGGPFTTGNNAAWYADNNGNLVKGFQ 1405
Qy 1365 -----TGARTINGOHLIFRA-NGVOVKGFEVTDHGRISYIDGNSGDO 1406
Db 1406 EIDGKWHYFDEVTGQQAALVNGQQLYFDVDSIQVKGFEVTDGQNGTSTSYDVSNGDK 1465
Qy 1407 INRRFVRNAQGWFFYFDNNGYAVTGARTINGOHLIF-RANGVOVKGFEVTDYGRISYVD 1465
Db 1466 KVGFEVTDGNNAAWYADGQGNLAKGRKSIDNODLIFDPATGQVKGQLVSD-GRNYIFD 1524
Qy 1466 ANSGERV 1473
Db 1525 SSGSGNMAK 1532

RESULT 15
Q8G9Q2 LEUME
ID Q8G9Q2 LEUME PRELIMINARY; PRT; 2835 AA.
AC Q8G9Q2.
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DI 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Dextranase (EC 2.4.1.5) (Fragment).
OS Name=darE;
GN Leuonostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuonostoc.
OX NCBI_TaxID=1245;
RN [1]
RX MEDLINE=22231661; PubMed=12270834;
RA Bozonnet S., Dols-Lafargue M., Fabre E., Pizut S., Renaud-Simeon M.,
RT "Molecular characterisation of DSR-E, an alpha-1,2 linkage
RT synthesising dextranase with two catalytic domains.";
RL J. Bacteriol. 184:5753-5761 (2002).
DR EMBL; AJ430204; CAD22883.1; -: Genomic DNA.
DR GO; GO:0047845; P:dextranase activity; IEA.
DR GO; GO:0006757; P:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 2.
KW Glycosyltransferase; Transferase.
FT NON_TER 1
SQ SEQUENCE 2835 AA; 313265 MW; D03262CDD735399D CRC64;

Query Match
Best Local Similarity 43.9%; Score 3398.5; DB 2; Length 2835;
Matches 706; Conservative 230; Mismatches 477; Indels 191; Gaps 22;

Qy 18 TVSVASAVMTLTTLTSGGLVKADSNESKQISNDSNTSVVTANEESNVITEATSKQEAASS 77
Db 121 TPAADKAVDTTPTPAADKAVDTTPTPAANKAVDTTPTAATDKAVATPAT---PAADK 177
Qy 78 QTNHTVTTSSSTSVNPKVSVNRYT-VGETASNGEKLQNTTVDTKTSEAAANNISKQ 136
Db 178 LANTTTPATD-----KAVATTPATPVANKAADTSIHDPQDNTNPTOKSANLVSTT 228
Qy 137 TTEADTDVIDSNAALQILEKLPVVKEDGKYVYDNGKVRTWFTLIADGKILHFDFT 196
Db 229 QKSTNQVQKSTETSHLQ-----EINGKTYFLDNGQVKKNFATIDGKVLFDKT 279
Qy 197 GAYTDTSDTWNKIDVTTRSNLYKKYQVDRSAQSFHVDHLYTAESWYRPKYILKDGK 256
Db 280 SGELETANAPQVTKGLV----NIDNAHAAHDLTADNFNVGDLTANSWYRPKDLKNGT 335
Qy 257 TWTQSTEKDFRPLMTWPPDQETQRYVYNNMAQIAGINKTYDDT-----SNOLQNLIAAA 311
Db 336 TWTPTTAEEDFRPLMSWPPDKNTQVAYLQYMQS---VGMPLPDVVKVSDNDNMSTLTDAA 392
Qy 312 TIQAKIEAKITTLKNTDMLRQTISAFVKTQSAWNSDSEKFPDDHLQNGAVLYDNEGKLTTP 371
```

```
Db 337 TVQKNIESRIGVSGKTDWLKQDMNKLIDSQANWNIDSESKGNHDLQGALLVYNDK-TP 451
Qy 372 YANSYRILNTPNTQCKDPRYTADNTIGGYEFLANDVNSNPVQAEOLNWLHFLAM 431
Db 452 NANSYRLNTPNTQCKDPRYTADNTIGGYEFLANDVNSNPVQAEOLNWLHFLAM 507
Qy 432 NFGNIYANDPANPDSIRVDAVNDVADLLOIAGDYKAAKGIHKNDKAAANDHLSILEAW 491
Db 508 NIGTIAQNDPTANFDGYRVDVNDVADLLOIAGDYKAAAYGTGKTEANNHHSILEDM 567
Qy 492 SDNDTPYLDDGDNMNMNKLRLSLFLAKPLNQRSGMNPPLIATNSLVNRDNDNAETAA 551
Db 568 DNDSDAYIKAHGNNQLTWDFFAHLALKYALNPLAAQSGLEPLINTSLVKRGKDATENEA 627
Qy 552 VPSYFTRAHDSEVDLIADIIKAEINPNVGVSYFTMEIEIKAFIYNKDLLATEKCYTH 611
Db 628 QPNTAFIRAHDSVQTVIAQIIKDKINTKSDGLTVPDEIKQAFIYNADDELKADKEYTA 687
Qy 612 YNTALSALLTNKSSVPRVYVGMFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAM 671
Db 688 YNIPASYAVLLTNKDTVPRVYVGDLSDDGQYMSQSPYYDAITSLLKSRIKYVAGGOSM 747
Qy 672 R-----NOQVGNSEIITSVRYCKGALKATDGTDRTRTTSVAVIEGNPNLSRLK 720
Db 748 NMTYLHECFDPAKNETKPGVLTSVRYCKGAMTADDLGNSTROQIGLIVNKKPFLNLN 807
Qy 721 ASDRVVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKG 780
Db 808 DDEQIVLNMGAHKNQAYRPLMLTTKSLQIY--DKDAGAPVYVYNDAGLIFKSDMVYV 865
Qy 781 YANQVSGYLGVWVPVGAALLKMPALRLARPHQMAVSV-HQNAALDSRVMEGPFNFQAF 839
Db 866 VSNPQVSGYFAAWPVVGASDSQARTQSSQSETKGDYVYHNAALDSNVIYEGFSNFQAM 925
Qy 840 ATKKEEYTNVVIKAVNDKFAEWGVTDFEMAPOYVSSTDCGSLDSVIQNGYAFTRDYDLGI 899
Db 926 PEKNDFFNVKIAQNAKLFKDLGITSFELAPQYRSSTDSNLSFLDSVIQNGYAFTRDYVGY 985
Qy 900 SKPNKYGTADDLVKAIKALHSKGIKVMADVPDQMAPEPEKEVVTATRVKYGTVPAGSQ 959
Db 986 NTPTKYGTVDQLDLSRLAHAQGIQAINDWVPDQIYNLPGEQIVTAVRTNGSGKYDYDSV 1045
Qy 960 IKNTLYVVDGKS-SGKDOQAKYGGAFLELOAKYPELFARKQISITGVPMDSVKIKQNSA 1018
Db 1046 INNTLY--DSRTVGGGEYQEXFPGGLFLDQLKKDYPSPLETKQISTQPNMNPVVKIKWSA 1103
Qy 1019 KYFNGTNILGRGAGYVLKDAQATNTVFNISDNKEINFLPKTLNQDSQVGSFDGKGYVY 1078
Db 1104 KYFNGSNIOGRGAWYVLKDWATNQYFNVSND--GFLPKQLLGEKTSSTGTENGKTSFY 1161
Qy 1079 STSGYQAKNTFISBGRKWIYFDNNGYVMTGAQSGINGVNYFLSGLQRLDAILKNDGTY 1138
Db 1162 STSGYQAKNTFISBGRKWIYFDNNGYVMTGAQSGINGVNYFLSGLQRLDAILKNDGTY 1221
Qy 1139 AYYGNDGRRYENGYYQPMGSGVWRH-FNNGEWSVGLTVI---DGO--VOYFDEMGOAKGK 1192
Db 1222 -YNNKAGEQVNNQYQDSQNGQWHYFFENGRMAIGLITEVPNADGTHVTQYFDANGVOIKGT 1280
Qy 1193 FVTTADGKIRYFDKQSGNMVNRNFTIENEEGKWLVLGEDCAAATGSGQTLNGOHLIFRANGV 1252
Db 1281 AIKQONNQLRFPDEATGNMNVNSWQLADKSWLYLNAQGVAVTGKQIDGEBEYFNADGK 1340
Qy 1253 QVKGFEVTDHGRISYVD----- 1270
Db 1341 QVKNALIIDNDGQRYDGDGKGMVMSWGBELPDGSWLYLNDKGIATVGTGRQVINNVNPF 1400
Qy 1271 GNSGQIIRNFRVRAQGWFFYFDNNGYAVTGARTINGOLLYFRANGVQVKGFEVTDYGR 1330
Db 1401 GNDGQIKDAPFKLLSDGSWVYLDKGLITTTAKVINGLNMFPFDKDHQIKGDASTDANGK 1460
Qy 1331 ISYVDGN-----SGQIRN----- 1344
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 136.863 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-35
Perfect score: 7143
Sequence: 1 MEKKVRFKLRKYKKRWTVS.....VKVSNSTMLIPMKFVIVM 1375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: geneseqp1980a:*
2: geneseqp1990a:*
3: geneseqp2000a:*
4: geneseqp2001a:*
5: geneseqp2002a:*
6: geneseqp2003a:*
7: geneseqp2003bs:*
8: geneseqp2004a:*
9: geneseqp2005a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7143	100.0	1375	5 AAU98028	Aau98028 S. mutans
2	7143	100.0	1375	5 AAU79288	Aau79288 Streptoco
3	7143	100.0	1375	7 ADD93655	Add93655 Streptoco
4	7143	100.0	1375	9 ADX37278	Adx37278 Streptoco
5	5392	75.5	1475	5 AAU79284	Aau79284 Streptoco
6	5285	74.0	1475	5 AAU98027	Aau98027 S. mutans
7	5285	74.0	1475	7 ADD93654	Add93654 Streptoco
8	5285	74.0	1475	9 ADX37277	Adx37277 Streptoco
9	5284	74.0	1475	5 AAU98030	Aau98030 S. mutans
10	5281	73.9	1475	5 AAU98040	Aau98040 S. mutans
11	5280	73.9	1475	5 AAU98031	Aau98031 S. mutans
12	5279	73.9	1475	5 AAU98035	Aau98035 S. mutans
13	5279	73.9	1475	5 AAU98033	Aau98033 S. mutans
14	5278	73.9	1475	5 AAU98032	Aau98032 S. mutans
15	5277	73.9	1475	5 AAU98036	Aau98036 S. mutans
16	5273	73.8	1475	5 AAU98034	Aau98034 S. mutans
17	5271	73.8	1475	5 AAU98037	Aau98037 S. mutans
18	5261	73.7	1475	5 AAU98038	Aau98038 S. mutans
19	5258	73.6	1475	5 AAU98039	Aau98039 S. mutans
20	4334.5	60.7	1017	5 AAU79285	Aau79285 Streptoco
21	4214.5	59.0	1590	7 ADD93657	Add93657 Streptoco
22	4214.5	59.0	1590	9 ADX37280	Adx37280 Streptoco
23	4168.5	58.4	1592	2 AAR32925	Aar32925 Glucosylt
24	3603.5	50.4	1430	5 AAU98044	Aau98044 S. mutans

ALIGNMENTS

RESULT 1
AAU98028
ID AAU98028 standard; protein; 1375 AA.
XX
AC AAU98028;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFC.
XX

KW Glucosyltransferase; GTFC; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture.
XX Streptococcus mutans.
XX
PN US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX

PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
PA (NICH//) NICHOLS S E.
XX
PI Nichols SE;
XX
XX WPI; 2002-414332/44.
XX N-PSDB; ABK52939.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture, comprises mutations in specific positions.

Disclosure; Page 30-33; 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF) B polypeptide having changes at position from I448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T, I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

Aau98045 S. mutans
Aau98041 S. mutans
Aau98043 S. mutans
Aau98042 S. mutans
Aau98029 S. mutans
Add93656 Streptoco
Adx37279 Streptoco
Adcs4806 Protein S
Aar91047 Alpha-D-g
Add93658 Streptoco
Adx37281 Streptoco
Abr63234 Glucanusc
Adcs4807 Leuconost
Aau80055 Leuconost
Abb98574 Dextran s
Abr55594 Amino aci
Add93660 Streptoco
Adx37283 Streptoco
Ady72696 Mutant de
Ady72733 Mutant de
Ady72732 Mutant de

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes F1 or F2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents GTFc
XX
SQ Sequence 1375 AA;

Query Match 100.0%; Score 7143; DB 5; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKKVRFKLRKVKRWVTSIASAVVTLTSLGSLVKADSTDDROQAVTESQASLVTTSE 60
DB 1 MEKKVRFKLRKVKRWVTSIASAVVTLTSLGSLVKADSTDDROQAVTESQASLVTTSE 60
QY 61 AAKETLTATDTSTATSAISQPTATVDNVSTNQSTNTTANTANFVVKPTTSEQAKTDN 120
DB 61 AAKETLTATDTSTATSAISQPTATVDNVSTNQSTNTTANTANFVVKPTTSEQAKTDN 120
QY 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIKPKIGKLKQPSLSODDIAA 180
DB 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIKPKIGKLKQPSLSODDIAA 180
QY 181 LGNVQKIRKNGKYYKEDGTLQKNYALNINGKTFPFDETGALSNNTLPSKKGNITNND 240
DB 181 LGNVKXIRKNGKYYKEDGTLQKNYALNINGKTFPFDETGALSNNTLPSKKGNITNND 240
QY 241 NTNSPAQYNQVSTDVANFEVDHVTAEBSWTRPKYILKDGKWTQSTOKDPRPLLLMTWW 300
DB 241 NTNSPAQYNQVSTDVANFEVDHVTAEBSWTRPKYILKDGKWTQSTOKDPRPLLLMTWW 300
QY 301 PDQETQROYVNMNAQLGHOTYNTATSPQLNLAAQTITQKIEKITAENKTNWLRQTI 360
DB 301 PDQETQROYVNMNAQLGHOTYNTATSPQLNLAAQTITQKIEKITAENKTNWLRQTI 360
QY 361 SAFVKTQSAWNSDSKPPDDHLQKALLYSNNSKULTSOANSNYRILNRTPTNQTGKQDPR 420
DB 361 SAFVKTQSAWNSDSKPPDDHLQKALLYSNNSKULTSOANSNYRILNRTPTNQTGKQDPR 420
QY 421 YTADRTTIGYEFLANDVNSNPVQAEQLNWLFLMNFNIYANDPDANFDSIRVDVAD 480
DB 421 YTADRTTIGYEFLANDVNSNPVQAEQLNWLFLMNFNIYANDPDANFDSIRVDVAD 480
QY 481 NVADLLQIAGDYLKAAKGIHKNDHLSILEAWSYNDTPPYLHDDGDNNINMNDNRLR 540
DB 481 NVADLLQIAGDYLKAAKGIHKNDHLSILEAWSYNDTPPYLHDDGDNNINMNDNRLR 540

QY 541 LSLLYSLAKPLNQRSGMNPPLITNSLVNRTDDNAETRAAVPSYFTRAHDSEVQDLIRNIIR 600
DB 541 LSLLYSLAKPLNQRSGMNPPLITNSLVNRTDDNAETRAAVPSYFTRAHDSEVQDLIRNIIR 600
QY 601 TEINPNVGVSYFTTEEEKAFEIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVG 660
DB 601 TEINPNVGVSYFTTEEEKAFEIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVG 660
QY 661 DMFTDDGQYMAHKTINVEAIEITLKARIKYVSGGQAMNQQVGNSEIITSVRYGKALK 720
DB 661 DMFTDDGQYMAHKTINVEAIEITLKARIKYVSGGQAMNQQVGNSEIITSVRYGKALK 720
QY 721 TDTGDRTRTSGVAVIEGNPSLRKASDRVVVNGAAHKNQAVRPLLLTTDNGIKAVHS 780
DB 721 TDTGDRTRTSGVAVIEGNPSLRKASDRVVVNGAAHKNQAVRPLLLTTDNGIKAVHS 780
QY 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVGVVPGAADQDVRVAASTAPST 840
DB 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLVGVVPGAADQDVRVAASTAPST 840
QY 841 DGKSVHQNAALDSRVMPFEGFSNFOAFATKBEYTNVVIKXNDVKPAEANGVTDFFEMAPQYV 900
DB 841 DGKSVHQNAALDSRVMPFEGFSNFOAFATKBEYTNVVIKXNDVKPAEANGVTDFFEMAPQYV 900
QY 901 SSTDGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVAIKALHSGIKVMADWVDDQ 960
DB 901 SSTDGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVAIKALHSGIKVMADWVDDQ 960
QY 961 MYALPEKEVTVATRVDKYGTVPAGSQIKNTLVYVDGSKSGKQQAQKYGAFLEELQAKYP 1020
DB 961 MYALPEKEVTVATRVDKYGTVPAGSQIKNTLVYVDGSKSGKQQAQKYGAFLEELQAKYP 1020
QY 1021 ELFPARKQISTGVPMDPPSVKIKQWSAKYPNGTNILGRGAGYVLKQQAQNTYFSLVSDNTFL 1080
DB 1021 ELFPARKQISTGVPMDPPSVKIKQWSAKYPNGTNILGRGAGYVLKQQAQNTYFSLVSDNTFL 1080
QY 1081 PKSLVNPNGHTSSSVTGLVFDGKGYVYSTSGNOAKNAFISLGNWYTFDNGYVMTGAQ 1140
DB 1081 PKSLVNPNGHTSSSVTGLVFDGKGYVYSTSGNOAKNAFISLGNWYTFDNGYVMTGAQ 1140
QY 1141 SINGANYFYLSNGIQLRNAIYDNGNKVLSYVYNGDGRYENGYYLFGQOWRYFQNGIMAVG 1200
DB 1141 SINGANYFYLSNGIQLRNAIYDNGNKVLSYVYNGDGRYENGYYLFGQOWRYFQNGIMAVG 1200
QY 1201 LTRVHGAVQYVFDASGFOAKGQFIITADGKLRYFDRDSGNQISNRPVRNSKGEWFLFDHNG 1260
DB 1201 LTRVHGAVQYVFDASGFOAKGQFIITADGKLRYFDRDSGNQISNRPVRNSKGEWFLFDHNG 1260
QY 1261 VAVTGTVTENGORLYEKPNQVQAKGEFTRDANGYLYRYDPSNGNEVRNRFVRNSKGEWFL 1320
DB 1261 VAVTGTVTENGORLYEKPNQVQAKGEFTRDANGYLYRYDPSNGNEVRNRFVRNSKGEWFL 1320
QY 1321 FDHNGIATGARVNVNGHASILSLMVFRLRESSLSQSVKVSNTMILIPMKFVIVM 1375
DB 1321 FDHNGIATGARVNVNGHASILSLMVFRLRESSLSQSVKVSNTMILIPMKFVIVM 1375
RESULT 2
AAU79288
ID AAU79288 standard; protein; 1375 AA.
XX AAU79288;
AC AAU79288;
XX
DT 13-AUG-2002 (first entry)
XX
DE Streptococcus mutans monoclonal antibody-related protein #5.
XX
KW Antibody; dental caries; water insoluble glucan synthetase; anti-caries;
XX glucosyl transferase-B; immunotherapy.
OS
XX Streptococcus mutans.

PN JP2002114709-A.
XX 16-APR-2002.
XX 04-OCT-2000; 2000JP-00304889.
XX PF 04-OCT-2000; 2000JP-00304889.
PR 04-OCT-2000; 2000JP-00304889.
XX (UYNI-) UNIV NIPPON.
XX WPI; 2002-4488q5/48.
DR Anti-carries agent composed of a monoclonal antibody against an inhibitory
PT enzyme against water insoluble glucan synthetase of glucosyl transferase-
PT B (GTF-B) of Streptococcus mutans.
XX Claim 3; Page 13-16; 28pp; Japanese.
XX The invention relates to a monoclonal antibody against dental caries and
CC an anti-carries agent composed of a monoclonal antibody produced by
CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)
CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
CC inhibitive activity against water insoluble glucan synthetase of glucosyl
CC transferase-B. The monoclonal antibody specifically inhibits water
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
CC transferase-B and is used in the immunotherapy of dental caries. This
CC sequence represents a Streptococcus mutans monoclonal antibody-related
CC protein
XX
SQ Sequence 1476 AA;

Query Match 75.5%; Score 5392; DB 5; Length 1476;
Best Local Similarity 77.6%; Pred. No. 0;
Matches 1044; Conservative 97; Mismatches 157; Indels 48; Gaps 11;

Qy 1 MEKVRFLKRLKVKKEWVTVASAVVTLTSLGSLVKADSTDDRQQAATESQASLVTTSE 60
Db 1 MDKRVYKRLKVKKRWVTVASAVVTLTSLGSLVKADSNESKSQISNDSNTSVVTANE 60

Qy 61 AAKETLATDTSTATSQPTA--TVTDNVSTTQSTNTTANTANFVVKPTTSEQAKT 118
Db 61 -----ESNVITEATSKQEAASSQTHVTTSSSTSVVNPKEVSNPYTVGETA-- 109

Qy 119 DNSKIIITTSKAVNKLRTATGFPVANNHTAHPKVTVDKIPIKPKIGLKPSSLSQDDI 178
Db 110 SNGEKLQNTTVDKTSEA----AANNISKQTTEAD-----TDVIDDSN 149

Qy 179 AA---LGNVKNIRKVVNGKYVYKEDGTLOKNYALNNGKTFPDETCALSNNTLPSKKG 234
Db 150 AANLQLEKLPVKEIDGKYVYDNGKVRNTFLIADGKILHFDGTGAYTDTSDITVKN 209

Qy 235 NITNDNTNSFAQYQVYSTDVANFEVDHYLTAEWSYRPRYILKDGKTWTQSTEDFRP 294
Db 210 DIVTT-RSNLYKYNQVYDRSAQSPFVDHYLTAEWSYRPRYILKDGKTWTQSTEDFRP 268

Qy 295 LLMTWPPQETQRYVNTMNAQLGHQTYNTATSPQLNLAQAOTIQTKEIKTAEKVTN 354
Db 269 LLMTWPPQETQRYVNTMNAQLGINKTYDDTNSQLNLNIAAATTAQAEAKITTLKNTD 328

Qy 355 WLRTTISAFVKTQSAWNSDEKPPDHLQKALLYSNNSKLTQSAWSNRYILNRTPTNQT 414
Db 329 WLRTTISAFVKTQSAWNSDEKPPDHLQKALLYSNNSKLTQSAWSNRYILNRTPTNQT 388

Qy 415 GKQPRYTADRTIGGYEFLANDVNSNPVQAEQLNLWHLFNMFGNIYANDPDPAFDSI 474
Db 389 GKQPRYTADRTIGGYEFLANDVNSNPVQAEQLNLWHLFNMFGNIYANDPDPAFDSI 448

Qy 475 RVDVNDVADLLQTAGDYLLKAAKGIHKNDKKAANDHLSILEAWSYNDTPYLHDGDNMIN 534
Db 449 RVDVNDVADLLQTAGDYLLKAAKGIHKNDKKAANDHLSILEAWSYNDTPYLHDGDNMIN 508

Qy 535 MDNRLRLSLYLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVPSYFIRAHDSEVQDL 594
XX

Db 509 MDNKLRLSLYLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVPSYFIRAHDSEVQDL 568
Qy 595 IRNIIRTEINPNVVGYSFTTEEIKKAFEIYNKDLLATEKKYTHYNTALSYALLLNKSSV 654
Db 569 IRDIIKABINPNVVGYSFTTEEIKKAFEIYNKDLLATEKKYTHYNTALSYALLLNKSSV 628
Qy 655 PRVYVGMFTDDGQYMAHKTIINYEAIETLLKARIKYVSGGQAMRNQOVNSEIITSVRYG 714
Db 629 PRVYVGMFTDDGQYMAHKTIINYEAIETLLKARIKYVSGGQAMRNQOVNSEIITSVRYG 688
Qy 715 KGALKATDGTDRTRTSGVAVIEGNNPSLRKASDRVVVNMGAHKNQAYRPLLLTTDNG 774
Db 689 KGALKATDGTDRTRTSGVAVIEGNNPSLRKASDRVVVNMGAHKNQAYRPLLLTTDNG 748
Qy 775 IKAYHSQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGVVMPVGAADQDVRVAA 834
Db 749 IKAYHSQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGVVMPVGAADQDVRVAA 808
Qy 835 STAPSTDGKSVHONALDSRVMPGFSNFQAFATKKEEYTNVIAKNYDKFAEWGVTDFE 894
Db 809 STAPSTDGKSVHONALDSRVMPGFSNFQAFATKKEEYTNVIAKNYDKFAEWGVTDFE 868
Qy 895 MAPOVVSSTGSLFDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKYMA 954
Db 869 MAPOVVSSTGSLFDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKYMA 928
Qy 955 DWVPDQMYALPEKEVVTATRVVDKYGTPVAGSQIKNTLVVDGKSGKQQAQKYGAFLEE 1014
Db 929 DWVPDQMYALPEKEVVTATRVVDKYGTPVAGSQIKNTLVVDGKSGKQQAQKYGAFLEE 988
Qy 1015 LOAKYPELPFARKQISTGVPMDPSPVKIKOWSAKYFNGTNILRGAGYVLKQATNTYPSLV 1074
Db 989 LOAKYPELPFARKQISTGVPMDPSPVKIKOWSAKYFNGTNILRGAGYVLKQATNTYFN-I 1047
Qy 1075 SDN---TFLPKSLVNPNGHTSSSVTGLVFDCKGVYVYSTSGNOAKNAFISLGNMYFDN 1131
Db 1048 SDNKEINFLPKTLN-----QDSQVGSYDCKGVYVYSTSGVQAKNTFISEGDKYVFDN 1102
Qy 1132 NGYVMTQAQSGINGANYFLSNGIQLRNAIYDNGKNVLSYVYDNGRRYENGYYLF-GQOWR 1190
Db 1103 NGYVMTQAQSGINGANYFLSNGIQLRDAILKNEIDGTAYAYDNGRRYENGYYQFMSGVR 1162
Qy 1191 YFQNGIMAVGLTRVHGAQVYFDASGFOAKGQFITTADKLRIFDRDSGNQISNRFVRNSK 1250
Db 1163 HFNNGEMSVGLTVIDGQVQYFDEMGOYAKGKFVTADGKIRYFDKQSGNMYRNFIEEE 1222
Qy 1251 GEWFLFDHNGVAVTGTTFNGORLYFKPNGVQAKGEFIRDANGYLRYVYDPSNGNEVRNRF 1310
Db 1223 GKWLVLGEDGARVTSQTINGQHLYFRANGVQVKGFEVTDHGRISYYDNGSGDQIRNRF 1282
Qy 1311 VRNSKGEWFLPDHNGIAVTVGARVNG 1336
Db 1283 VRNAQGWYFDDNGYAVTGARTING 1308

RESULT 6
AAU98027
ID AAU98027 standard; protein; 1475 AA.
XX
AC AAU98027;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFB.
XX
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
XX amyloplast; vacuole; paper manufacture.
XX Streptococcus mutans.
XX
PN US2002031826-A1.
XX

QY 1189 WRYFONGIMAVCLTRVHGAVQYFDASGFOAKGQFITTADGKLRYPDRDSGNQISNRFVRN 1248
 DB 1160 WRHFNNGEMSVGLTVIDGVOYFDEMGYQAKGKFVTTADGKIRYPDKSGNMRYNRFTEN 1219
 QY 1249 SKGEWFLPDHNGVAVTGVTFNGORLYKPKNGVQAKGFIIRDANGYLRYYPNSGNEVRN 1308
 DB 1220 BEGKWLXIGEDGAAVTGSGTNGQHLVFRANGVQVKGSEFVTDHHRISYDGNSGDQIRN 1279
 QY 1309 RFVRNSKGWFLFDHNGIAVTCARVVGNGHASILSLMVR 1347
 DB 1280 RFVRNAQGWFFFDNNGVAVTGARTING-----QLLYFR 1313

RESULT 7

ADD93654

ID ADD93654 standard; protein; 1475 AA.

AC ADD93654;

29-JAN-2004 (first entry)

Streptococcus mutans glucosyltransferase-B.

Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

Streptococcus mutans.

WO2003075845-A2.

18-SEP-2003. 1

07-MAR-2003; 2003WO-US006962.

07-MAR-2002; 2002US-0363209P.

08-AUG-2002; 2002US-0402483P.

(FORS-) FORSYTH INST.

Smith DJ, Taubman MA;

WPI; 2003-845091/78.

Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

Claim 16; Page 12-13; 49pp; English.

The present sequence is the protein sequence of *Streptococcus mutans* glucosyltransferase-B (GTF-B). Peptide fragments of GTF-B, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from *S. mutans* glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multicitepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

Sequence 1475 AA;

Query Match 74.0%; Score 5285; DB 7; Length 1475;

Best Local Similarity 75.9%; Pred. No. 0;

Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;

QY 1 MEKKVRFKLRYKRWVTVSIAVVTLTSLSGSLVKADSTDRQQAUTESQASLVTTSE 60

DB 1 MDKVRVYKLRVKRWVTVSIVASAVMTLTSLGGLVKADSNESQISNDSNTSVVTANE 60

QY 61 AAKEILTATDTSATISATSQPTA--TVTDNVSTNQSTNTTANTANFVVVKPTTSEQAKT 118

DB 61 -----ESNVITEATSKOEAASSQTNHTVTTSSSTSVVNPKEVSVNPVTVGETA-- 109
 QY 119 DNSDKIITTSKAVNRLTATGRFVPPANNNTAHPKVTVDKIVIPKIGKLGKOPSSLSQDDI 178
 DB 110 SNGEKLQNTTVDKTSSEA-----AANWISQOTTEAD-----TDVIDDSN 149
 QY 179 AA-----LGNVKNIRKNGKYVYKEDGTLQKNYALNINGKTFPFDETGALSNTTLPKKG 234
 DB 150 AANLQILEKLPNVKEIDGKYVYDNGKVRNFTLIADGKILHFDDETGAYDTSIDTVNK 209
 QY 235 NITNDNTNSFAOYNQVSTVDANFEHVDHYLTAEWSWRPKYILKDGKTWTQSTEKDFRP 294
 DB 210 DIVTT-RSNLYKKNQVYDRSAQSFEHVDHYLTAEWSWRPKYILKDGKTWTQSTEKDFRP 268
 QY 295 LLMTWMPDQETQROYVNTMNAQLGIHQTYNTATSPQLNLAAQTITQKIEEKITAENKTN 354
 DB 269 LLMTWMPDQETQROYVNTMNAQLGINKTYDDTSNQLNLAAATIAQAKIETTLTKNTD 328
 QY 355 WLQRTISAFVKTQSAWNSDSEKPFDDHLQKALLYNNNSKLTQSAANSYRILNRTPTNQ 414
 DB 329 WLQRTISAFVKTQSAWNSDSEKPFDDHLQNGAVLYDNEGKLTTPYANSYRILNRTPTNQ 388
 QY 415 GKQDPRTYADRTIGGYEFLLANDVDSNPVQAEQLNLWLFHLMNFGNIYANDPDANFDSI 474
 DB 389 GKQDPRTYADRTIGGYEFLLANDVDSNPVQAEQLNLWLFHLMNFGNIYANDPDANFDSI 448
 QY 475 RVDADVNDADLLQIAGDYLKAAKGIIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNMIN 534
 DB 449 RVDADVNDADLLQIAGDYLKAAKGIIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNMIN 508
 QY 535 MDNRLRLSLYSLAKPLNQRSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDL 594
 DB 509 MDNKLRLSLFLAKPLNQRSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDL 568
 QY 595 IRNIIRTEINPNVVGYSPTTEEIKKAPRIYKNDLLATEKKYTHYNTALSYALLTNKSSV 654
 DB 569 IADIIKAEINPNVVGYSFTMBEIKKAFRIYKNDLLATEKKYTHYNTALSYALLTNKSSV 628
 QY 655 PRVYVYGDMDTDDGOYMAHKTINYBAIETLLKARIKYVSGGQAMRNQOVGNSEIITSVRYG 714
 DB 629 PRVYVYGDMDTDDGOYMAHKTINYBAIETLLKARIKYVSGGQAMRNQOVGNSEIITSVRYG 688
 QY 715 KGALKATDTGDRTRTSGVAVIEGNPNPSRLKASDRVVVNMGAHKNQAYRPLLLTTDNG 774
 DB 689 KGALKATDTGDRTRTSGVAVIEGNPNPSRLKASDRVVVNMGAHKNQAYRPLLLTTDNG 748
 QY 775 IKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYNAPQVSGYLVWVPVCAAADQ--DVRV 832
 DB 749 IKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYNAPQVSGYLVWVPVCAALIKKFAURL 808
 QY 833 AASTAPSTDGKSVHQNAALDSRVNPEGSFNFQAFATKKEEYTNVVIKKNVDKFAEWGVT 892
 DB 809 A---RPHQQMASVHQNAALDSRVNPEGSFNFQAFATKKEEYTNVVIKKNVDKFAEWGVT 865
 QY 893 FEMAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGIKV 952
 DB 866 FEMAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGIKV 925
 QY 953 MADVVPDQMYALPEKEVVVTRVDKYGTPVAGSQIKNTLYVVDGKSSGKQDQAKYGGAPL 1012
 DB 926 MADVVPDQMYALPEKEVVVTRVDKYGTPVAGSQIKNTLYVVDGKSSGKQDQAKYGGAPL 985
 QY 1013 EELQAKYFELFARKQISTGVPMDSVKIKQWSAKYFNGTNIILGRGAGVYLKQDQATNTYPS 1072
 DB 986 EELQAKYFELFARKQISTGVPMDSVKIKQWSAKYFNGTNIILGRGAGVYLKQDQATNTYFN 1045
 QY 1073 LVSDN---TFLPKSLVNPNGHTSSVTGLVPFGKGYVYVYTSNGQAKNAFISLGNWYIF 1129
 DB 1046 -ISDNKEINFLPKTLN-----QDSQVGFSDGKGYVYVYTSNGYQAKNTFISEGDKWYIF 1099
 QY 1130 DNNGYVMTGAGSINGANYFELSNGLQRLNLYDNGNKVLSYVYDNGDRYENGYILF-GQQ 1188

QY 1130 DNNGYMTGAOSINGANYFFLSNGIQLRNAYIDNGKNVLSYVNGDRYENGYYLF -GQQ 1188
D5 1100 DNNGYMTGAOSINGANYFFLSNGIQLRDALIKNEDGYAYVNGDRYENGYYFMSGV 1159
QY 1189 WRYFQNGIMAVGLTRVHGAVQYFDASGFAKQGFITTAGDKLRYFDRDSGNQISNRFPVRN 1248
D5 1160 WRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNFIE 1219
QY 1249 SKGEWFLFDHNGVATGVTTFNGQRLYFKPQNGVQAKGEFIRDANGLYLRYDPSNGNEVRN 1308
D5 1220 EEGKWLVLGEDAAVQTGQINGOHLVFRANGVQVKGFEVTDHHRISYYDNGSDQIRN 1279
QY 1309 RFRVRSKGEWFLFDHNGVATGVTTFNGQRLYFKPQNGVQAKGEFIRDANGLYLRYDPSNGNEVRN 1347
D5 1280 RFRVRSKGEWFLFDHNGVATGVTTFNGQRLYFKPQNGVQAKGEFIRDANGLYLRYDPSNGNEVRN 1313

RESULT 9
AAU98030
ID AAU98030 standard; protein; 1475 AA.
XX
AC AAU98030;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFB mutant I448V.
XX
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.

XX
FH Key Location/Qualifiers
FT Misc-difference 448 /note= "Wild-type Ile substituted by Val"
XX
XX US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
PA (NICH/) NICHOLS S E.
XX
PI Nichols SE;
XX
DR WPI; 2002-414332/44.
XX

XX
PT Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
XX Claim 36; Page; 44pp; English.

XX
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from I589D, T589E, N471D,
CC N471D/T589E, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises and
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX
SQ Sequence 1475 AA;

Query Match 74.0%; Score 5284; DB 5; Length 1475;
Best Local Similarity 75.8%; Pred. No. 0;
Matches 1030; Conservative 102; Mismatches 169; Indels 58; Gaps 14;
QY 1 MEKVRPKLRKVRKRWTVSTASAVVTLTSLSGSLVAKADSTDDROQAVTESQASLVTTSE 60
D5 1 MDKVRKLRKVRKRWTVSVASAVVTLTSLSGSLVAKADSTDDROQAVTESQASLVTTSE 60
QY 61 AAKETLTATDTSTATSQTPTA--TVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKT 118
D5 61 -----ESNVITEATSKQEAASSQTNHTVTSSTSSVTVNPKVSNPVTGETA-- 109
QY 119 DNSDKIITTSKAVNRLTATGKFPVANNNTAPKVTVDKIVPKPKIGLKQPSLSDDI 178
D5 110 SNGEKLQQTTVTDKTSQA-----AANNISKQTTEAD-----TDVDDSN 149
QY 179 AA---LGNVKNIRKVGKYYYKEDGTLOKNYALNINIKTFFDETGALSNTLPSKKG 234
D5 150 AANLQILEKLPNVKEIDKGYYYDNGKVRNFTLIADGKILHFDGTGAYTDSIDTVNK 209
QY 235 NITNDNTNSFAQYQVYSTDVANFPHVDHYLTAEWSYRPKYILKDGKTWTQSTKDPFP 294
D5 210 DIVTT-RSNLYKYNQVYDRSAQSFHVDHYLTAEWSYRPKYILKDGKTWTQSTKDPFP 268
QY 295 LLMTWVDPQETQRYVNTMNAQLGHIQTYNTATSPQLNLAATQTKIEEKITAENYN 354
D5 269 LLMTWVDPQETQRYVNTMNAQLGHIQTYNTATSPQLNLAATQTKIEEKITAENYN 328
QY 355 WLRTTISAFVKTQSAWNSDSEKPPDDHLQKALLYSNNKSLTSQANSVNRILNRPPTNQ 414
D5 329 WLRTTISAFVKTQSAWNSDSEKPPDDHLQKALLYSNNKSLTSQANSVNRILNRPPTNQ 388
QY 415 GKCPRYTADRTTIGGYEFLANDVNSNPVQAEQLNWLHFLMNFNGIYANDPDANFDSI 474
D5 389 GKCPRYTADRTTIGGYEFLANDVNSNPVQAEQLNWLHFLMNFNGIYANDPDANFDSV 448
QY 475 RVDADVNDADLLQIAGYILKAAGIKHNDKAANDHLSILEAWSYNDTPYLLHDDGNM 534
D5 449 RVDADVNDADLLQIAGYILKAAGIKHNDKAANDHLSILEAWSYNDTPYLLHDDGNM 508
QY 535 MDNRLRLSLLYSLAKPLNQRSGMNPITNSLVNRVTDNAETAAPVSPYSFIRAHDSVQDL 594

Db 509 MNKRLSLFLSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPSPYSPIRADSEVQDL 568
QY 595 IRNITRTENPNVVGYSFTTEIKKAFELYNKDLATEKKYTHYNTALSVALLTNKSSV 654
Db 569 IADIIKAEINPNVVGYSFTTEIKKAFELYNKDLATEKKYTHYNTALSYALLTNKSSV 628
QY 655 PRVYGDMDTDDQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVNSEIITSVRYG 714
Db 629 PRVYGDMDTDDQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVNSEIITSVRYG 688
QY 715 KGALKATDGTTRTSGVAVTEGNPNPSRLKASDRVVVNMGAHKNAQYRPLLLTTDNG 774
Db 689 KGALKATDGTTRTSGVAVTEGNPNPSRLKASDRVVVNMGAHKNAQYRPLLLTTDNG 748
QY 775 IKAYHSDQEAAGLVRYTNRDGBELIFTAADIKGYANPQVSGYLGWVPVGVGAADO--DVRV 832
Db 749 IKAYHSDQEAAGLVRYTNRDGBELIFTAADIKGYANPQVSGYLGWVPVGVGAADO--DVRV 808
QY 833 AASTAPSTDGKSHVQNAALDSRVMEFGFSNFQAFATKKEEYTNVVIKNDKFAEWGVTD 892
Db 809 A--RPHQQWASVHQAALDSRVMEFGFSNFQAFATKKEEYTNVVIKNDKFAEWGVTD 865
QY 893 FEMAPQYVSTSGFLSDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGKIKV 952
Db 866 FEMAPQYVSTSGFLSDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGKIKV 925
QY 953 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGSSGKQOAKYGGAPL 1012
Db 926 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGSSGKQOAKYGGAPL 985
QY 1013 EELQAKYPELFAKQISTGVPMDPSPVKIKQWAKYFNGTNIILGRGAGYVLKQDQATNTVFS 1072
Db 986 EELQAKYPELFAKQISTGVPMDPSPVKIKQWAKYFNGTNIILGRGAGYVLKQDQATNTVFS 1045
QY 1073 LVSDN---TFLPKSLVNPNGHTSSVTGLVFDGKGVYVYSTSGNQAQNAFISLGNWYVF 1129
Db 1046 -ISDNKEINFLPKTLN-----QDSQVGFSDYDGKGVYVYSTSGYQAKNTFISEGDKWYVF 1099
QY 1130 DNNGYMVTGAQSGINGANYFLSNGQLRQNAIYDNGKNVLSYXGNDGRVYENGYYLP-CQQ 1188
Db 1100 DNNGYMVTGAQSGINGANYFLSNGQLRQNAIYDNGKNVLSYXGNDGRVYENGYYLP-CQQ 1159
QY 1189 WRYFQNGIMAVGLTRVHGAVQYFDASGFOAKGQFITTADGKLYRFDROSGNQISNRVRN 1248
Db 1160 WRFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTTADGKLYRFDROSGNQISNRVRN 1219
QY 1249 SKGEWFLFDHNGVAVTGTTFNGQRLYFKPVGQAKGFIIRDANGLYRYDPNSGNEVRN 1308
Db 1220 BEGKWLXLYLGEDGAAVTGSQTNGHLYFRANGVQVKGFEVTDHGRISYDNGSGDQIRN 1279
QY 1309 RFRVRSKGEWFLFDHNGVAVTGTTFNGQRLYFKPVGQAKGFIIRDANGLYRYDPNSGNEVRN 1347
Db 1280 RFRVRSKGEWFLFDHNGVAVTGTTFNGQRLYFKPVGQAKGFIIRDANGLYRYDPNSGNEVRN 1313

RESULT 10

ID AAU98040 standard; protein; 1475 AA.
XX
AC AAU98040;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFB mutant K779Q.
XX
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
XX coating composition; glucan; starch; latex; thermoplastic molecule;
XX amyloplast; vacuole; paper manufacture; mutant; mutin.
XX
OS Streptococcus mutans.
XX Synthetic.

FH Key Location/Qualifiers

FT Misc-difference 779 /note= "Wild-type Lys substituted by Gln"
XX
PN US2002031826-A1.
XX 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
DR
XX Glucosyltransferase B or D protein useful for producing a glucan useful
as substitutes for and additions to modified starch and latexes in paper
manufacture, comprises mutations in specific positions.
XX Claim 36; Page; 44pp; English.
XX
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
B polypeptide having changes at position from 1448V, D457N, D567T,
K1014T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
K148V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
GTF D polypeptide having changes at positions from T598D, T598E, N471D,
N471D/T598D, and N471D/T598E. Also included are a glucan produced by the
GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
complementary polynucleotide, a ribonucleic acid sequence encoding the
GTF mutant, an expression cassette comprising the polynucleotide operably
linked to a promoter, a vector comprising the expression cassette, host
cell introduced with the vector, a transgenic plant comprising the
vector, a seed or tuber from the transgenic plant, a paper sizing and/or
coating composition comprising a glucan produced in a plant transformed
with a gene encoding the mutant GTF, wild type or, starch, a latex,
thermoplastic molecule or their combinations or glucan and starch where
the glucan is produced in the amyloplast and/or vacuole or a maize line
deficient in starch biosynthesis, transformed with a gene encoding a
glucosyltransferase B or D enzyme, wild-type or mutant and a paper
comprising the glucan (paper sizing/coating agent). The vector is useful
for producing a glucan in a plant. The method comprises transforming a
plant cell with the vector, growing the plant cell under plant growing
conditions to produce a regenerated plant and inducing expression of the
polynucleotide for a time sufficient to produce the glucan in the
regenerated plant, where the vector contains a transit sequence from
ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
chlorophyll AB binding protein to produce a transgenic plant, and glucan
is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
and latexes in paper manufacture. Unlike prior art techniques, which
require input materials that produce chemical effluents, paper
manufacture utilizing the glucan produced by GTF, which utilizes
biologically produced input materials, is more cost-effective and
environmentally friendly. Moreover, glucans also exhibit thermoplastic
properties and impart gloss to the paper during coating step. The present
sequence represents a GTFB mutant of the invention. Note: The present
sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
claim 36
XX Sequence 1475 AA;
SQ Query Match 73.9%; Score 5281; DB 5; Length 1475;
Best Local Similarity 75.8%; Pred. No. 0;

Db 329 WLRQTISAFVKTSQAWNSDESKPDDHLQNGAVLYDNEGKLTTPYANSYRILNRPPTNQ 388
Qy 415 GKDPRTYADRTIGGYEFLANDVNSVPVQAEQLANLWHLFNLNFGNIYANDPDANFOSI 474
Db 389 GKDPRTYADRTIGGYEFLANDVNSVPVQAEQLANLWHLFNLNFGNIYANDPDANFOSI 448
Qy 475 RYDAVDNDADLLQTAGDYLAAGIKHNDKANDHLSILEAWSYNDTPYLLHDDGDNMIN 534
Db 449 RYDAVDNDADLLQTAGDYLAAGIKHNDKANDHLSILEAWSYNDTPYLLHDDGDNMIN 508
Qy 535 MDNRLSLYSLAKPLNORSQGNPLITNSLVRNTRDDNAETAAPVSPYSFIRAHDSQVODL 594
Db 509 MDNRLSLYSLAKPLNORSQGNPLITNSLVRNTRDDNAETAAPVSPYSFIRAHDSQVODL 568
Qy 595 IRNIIRTEINPNVVGYSFTTEIEKKAPEIYNKDLATEKKYTHYNTALSYALLLNKSSV 654
Db 569 IADIIKABINPNVVGYSFTTEIEKKAPEIYNKDLATEKKYTHYNTALSYALLLNKSSV 628
Qy 655 PRVYVYGDPTDDQYMAHKTINYEAIETLLKARIKYVSGGQAMRNOQVGNSEIITSVRYG 714
Db 629 PRVYVYGDPTDDQYMAHKTINYEAIETLLKARIKYVSGGQAMRNOQVGNSEIITSVRYG 688
Qy 715 KGALKATDTGRTTTSVAVTEGNNPSRLKASDRVVVNMGAHKNQAYRPLLLTTDNG 774
Db 689 KGALKATDTGRTTTSVAVTEGNNPSRLKASDRVVVNMGAHKNQAYRPLLLTTDNG 748
Qy 775 IKAYHSDQEAAGLVRYTNRGELIFTAADIKGYANPQVSGYLVGVVVPVGAADQ--DVRV 832
Db 749 IKAYHSDQEAAGLVRYTNRGELIFTAADIKGYANPQVSGYLVGVVVPVGAALIKMFALRL 808
Qy 833 AASTARSTDGKSVHQAALDSRWMEFSGNSFOAFATKKEEYTNVVIKXNDKPAEWGTYD 892
Db 809 A---RPHQOMASVHQAALDSRWMEFSGNSFOAFATKKEEYTNVVIKXNDKPAEWGTYD 865
Qy 893 FEMAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKV 952
Db 866 FEMAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKV 925
Qy 953 MADWVPDQMYALPEKEVVTATRVDKYTPVAGSQIKNTLYYVDGKSGKQOQAKYGGAPL 1012
Db 926 MADWVPDQMYAPPEKEVVTATRVDKYTPVAGSQIKNTLYYVDGKSGKQOQAKYGGAPL 985
Qy 1013 BELQAKYBELFARKQISTGVPMDPSPVKIKQNSAKYFNGTILGRGAGYLVKQDQATNTYFS 1072
Db 986 BELQAKYBELFARKQISTGVPMDPSPVKIKQNSAKYFNGTILGRGAGYLVKQDQATNTYFS 1045
Qy 1073 LVSDN---TFLPKSLVNPNGHTSSVTGLVPDGKGYVYVYSTSGNOKNAFTSLGNWYTF 1129
Db 1046 -ISDNKEINFLPRLN-----QDSQVGSFYDGKGYVYVYSTSGYQAKNFTFISEGDKWYTF 1099
Qy 1130 DNNGYMTVGAQSGINGANYFYFLNSGILQRLNAYIDNGNKNVLSYYGNDGRYENGYYLF-GQQ 1188
Db 1100 DNNGYMTVGAQSGINGANYFYFLNSGILQRLDAILKNEDGYVAYYNGDGRYENGYYQFMSGV 1159
Qy 1189 WRYFONGIMAVGLTRVHGAVQFDASGQAKQFITTADGKURYPDRSGNQISNRFRVN 1248
Db 1160 WRHFNNGEMSVGLTVIDGQVQYFDEMGYQAKGKFTTADGKIRYFDPKQSGNNRYNRFTEN 1219
Qy 1249 SKGEWFLFDHNGVAVTGTVTENGORLYPKPNVGQAKGFEIRDANGLYRYDPSNGNEVRN 1308
Db 1220 EGGKWLVLGEDAAVATGSGQTINGQHLFYFRANGVQVKGEFVTDHHRISYIDNGSGDQIRN 1279
Qy 1309 RFRNRSGEWFLLDHNGLAVTGARVNVNGHASTLSLWFR 1347
Db 1280 RFRNQAQGWFFDNDNGVAVTGCARTING-----QLLYFR 1313

RESULT 14
AAU98032
ID AAU98032 standard; protein; 1475 AA.
XX
AC AAU98032;

XX 27-AUG-2002 (first entry)
DT XX
DE S. mutans glucosyltransferase GTFB mutant D567T.
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutin.
XX
OS Streptococcus mutans.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
FT
FN US2002031826-A1.
PN 14-MAR-2002.
PD
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
PI
DR WPI; 2002-414332/44.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
as substitutes for and additions to modified starch and latexes in paper
manufacture, comprises mutations in specific positions.
XX Claim 36; Page: 44pp; English.
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
B polypeptide having changes at position from I448V, D457N, D567T,
K1014T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a
GTF D polypeptide having changes at positions from T589D, T589E, N471D,
N471D/T589D, and I471D/T589E. Also included are a glucan produced by the
GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
complementary polynucleotide, a ribonucleic acid sequence encoding the
GTF mutant, an expression cassette comprising the polynucleotide operably
linked to a promoter, a vector comprising the expression cassette, host
cell introduced with the vector, a transgenic plant comprising the
vector, a seed or tuber from the transgenic plant, a paper sizing and/or
coating composition comprising a glucan produced in a plant transformed
with a gene encoding the mutant GTF, wild type or, starch, a latex,
thermoplastic molecule or their combinations or glucan and starch where
the glucan is produced in the amyloplast and/or vacuole or a maize line
deficient in starch biosynthesis, transformed with a gene encoding a
glucosyltransferase B or D enzyme, wild-type or mutant and a paper
comprising the glucan (paper sizing/coating agent). The vector is useful
for producing a glucan in a plant. The method comprises transforming a
plant cell with the vector, growing the plant cell under plant growing
conditions to produce a regenerated plant and inducing expression of the
polynucleotide for a time sufficient to produce the glucan in the
regenerated plant, where the vector contains a transit sequence from
ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
chlorophyll AB binding protein to produce a transgenic plant, and glucan
is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
and latexes in paper manufacture. Unlike prior art techniques, which
require input materials that produce chemical effluents, paper
manufacture utilising the glucan produced by GTF, which utilises

Search completed: February 11, 2006, 19:14:42
Job time : 141.863 secs

1998 Feb 14 19:14:42

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Cpopyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 30.1046 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-35
Perfect score: 7143
Sequence: 1 MEKKVRFKLKRVKRWVTS.....VKVSNMILIPMKFVIM 1375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: +
1: /cgn2_6/prodata/1/iaa/5 COMB.pep: +
2: /cgn2_6/prodata/1/iaa/6 COMB.pep: +
3: /cgn2_6/prodata/1/iaa/H COMB.pep: +
4: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep: +
5: /cgn2_6/prodata/1/iaa/RE COMB.pep: +
6: /cgn2_6/prodata/1/iaa/backfiles.pep: +

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7143	100.0	1375	2	US-09-210-361-4
2	7143	100.0	1375	2	US-09-740-274-4
3	5285	74.0	1475	2	US-09-007-999-2
4	5285	74.0	1475	2	US-09-210-361-2
5	5285	74.0	1475	2	US-09-740-274-2
6	3591.5	50.3	1430	2	US-09-008-172-2
7	3591.5	50.3	1430	2	US-09-210-361-6
8	3591.5	50.3	1430	2	US-09-740-274-6
9	3216	45.0	1577	1	US-08-793-824-2
10	2501.5	35.0	2057	2	US-09-499-203-2
11	2347.5	32.9	1278	2	US-09-604-957-3
12	2347.5	32.9	1781	2	US-09-995-749A-2
13	1787	25.0	349	2	US-09-009-620-2
14	1718.5	24.1	545	2	US-09-604-957-4
15	1690.5	23.7	545	2	US-09-995-749A-10
16	1620.5	22.7	523	2	US-09-604-957-5
17	1616	22.6	522	2	US-09-995-749A-11
18	1445.5	20.2	535	2	US-09-604-957-7
19	1445.5	20.2	535	2	US-09-995-749A-13
20	1337	18.7	584	2	US-09-995-749A-12
21	1333	18.7	584	2	US-09-604-957-6
22	312.5	4.4	2710	1	US-08-480-604A-6
23	312.5	4.4	2710	1	US-08-405-496A-6
24	312.5	4.4	2710	2	US-08-915-136-6
25	312.5	4.4	2710	2	US-08-957-310-6
26	312.5	4.4	2710	2	US-10-011-366-6
27	312.5	4.4	2710	2	US-09-084-517-6

28	272	3.8	1742	2	US-09-386-962C-4	Sequence 4, Appli
29	272	3.8	1742	2	US-09-386-959-4	Sequence 4, Appli
30	269.5	3.8	811	1	US-08-480-604A-7	Sequence 7, Appli
31	269.5	3.8	811	1	US-08-405-496A-7	Sequence 7, Appli
32	269.5	3.8	811	2	US-08-915-136-7	Sequence 7, Appli
33	269.5	3.8	811	2	US-08-957-310-7	Sequence 7, Appli
34	269.5	3.8	811	2	US-10-011-366-7	Sequence 7, Appli
35	269.5	3.8	811	2	US-09-084-517-7	Sequence 7, Appli
36	269.5	3.8	812	1	US-08-480-604A-29	Sequence 29, Appli
37	269.5	3.8	812	2	US-08-915-136-29	Sequence 29, Appli
38	269.5	3.8	812	2	US-09-084-517-29	Sequence 29, Appli
39	268.5	3.8	2366	1	US-08-480-604A-10	Sequence 10, Appli
40	268.5	3.8	2366	1	US-08-405-496A-10	Sequence 10, Appli
41	268.5	3.8	2366	2	US-08-915-136-10	Sequence 10, Appli
42	268.5	3.8	2366	2	US-08-957-310-10	Sequence 10, Appli
43	268.5	3.8	2366	2	US-10-011-366-10	Sequence 10, Appli
44	268.5	3.8	2366	2	US-09-084-517-10	Sequence 10, Appli
45	266	3.7	1155	2	US-09-710-279-1780	Sequence 1780, Ap

ALIGNMENTS

RESULT 1
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; TYPE: PRT
; LENGTH: 1375
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match	100.0%;	Score 7143;	DB 2;	Length 1375;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1375;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MEKKVRFKLKRVKRWVTSIASAVVTLTSLGSLVKADSTDDRQAVTESQASLVTTS	60	
Db	1	MEKKVRFKLKRVKRWVTSIASAVVTLTSLGSLVKADSTDDRQAVTESQASLVTTS	60	
QY	61	AAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKTDN	120	
Db	61	AAKETLTATDTSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKTDN	120	
QY	121	SDKIITTSKAVNRLTATGKVFANNNHTAHPKVTVDKIVIPKIGLKOPSSLSQDDIAA	180	
Db	121	SDKIITTSKAVNRLTATGKVFANNNHTAHPKVTVDKIVIPKIGLKOPSSLSQDDIAA	180	
QY	181	LGNVKNIRKNGKYYYKEDGTGLQKNYALNINNGKTFEFTGALSNNITLPSKKGNTNND	240	
Db	181	LGNVKNIRKNGKYYYKEDGTGLQKNYALNINNGKTFEFTGALSNNITLPSKKGNTNND	240	

```
QY 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTOKDPRPLMTWM 300
Db 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTOKDPRPLMTWM 300

QY 301 PQDETQRYVYNNNAQLGHIQTYNTATSPQLNLAQAQTIQTKIEEKITAENKTNWLRTI 360
Db 301 PQDETQRYVYNNNAQLGHIQTYNTATSPQLNLAQAQTIQTKIEEKITAENKTNWLRTI 360

QY 361 SAFVKTQSAWNSSEKPFDDHLQKGLLYSNNKLTQSAWSNYRILNRTPTNQTGKQDPR 420
Db 361 SAFVKTQSAWNSSEKPFDDHLQKGLLYSNNKLTQSAWSNYRILNRTPTNQTGKQDPR 420

QY 421 YTDARTIGGYEFLFLLANDVNSNPVQAEOLNHLFLLMNFNFIYANDPDANFOSIRVDAVD 480
Db 421 YTDARTIGGYEFLFLLANDVNSNPVQAEOLNHLFLLMNFNFIYANDPDANFOSIRVDAVD 480

QY 481 NVDAADLLQAGDYLKAAGTGHKNDKAANDHLSILEAWSYNDTPYLLHDDGDNMINMDNRLR 540
Db 481 NVDAADLLQAGDYLKAAGTGHKNDKAANDHLSILEAWSYNDTPYLLHDDGDNMINMDNRLR 540

QY 541 LSLLYSLAKPLNORSQGNPLITNSLVNRTDDNAETAAPVSPYSFIRAHDSQVQDLIRNIIR 600
Db 541 LSLLYSLAKPLNORSQGNPLITNSLVNRTDDNAETAAPVSPYSFIRAHDSQVQDLIRNIIR 600

QY 601 TEINPNVGYSPFTEETEEIKAFELIYNKDLATEKKYTHYNTALSYALLTNKSVPRVYVG 660
Db 601 TEINPNVGYSPFTEETEEIKAFELIYNKDLATEKKYTHYNTALSYALLTNKSVPRVYVG 660

QY 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSIIITSVRYGKALKA 720
Db 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNQOVGNSIIITSVRYGKALKA 720

QY 721 TDTGDRTRTSGVAVIEGNPNLSRLKASDRVVVNMGAHKNQAYRPLLLTTTNGIKAYHS 780
Db 721 TDTGDRTRTSGVAVIEGNPNLSRLKASDRVVVNMGAHKNQAYRPLLLTTTNGIKAYHS 780

QY 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASAPST 840
Db 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASAPST 840

QY 841 DGKSVHQNAAALSRVNFEGSFNFQAFATKKEEYTNVVIKNDVKPAEWGVTDFEMAPQVY 900
Db 841 DGKSVHQNAAALSRVNFEGSFNFQAFATKKEEYTNVVIKNDVKPAEWGVTDFEMAPQVY 900

QY 901 SSTDGSFLDSVIONGVAFTRDYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
Db 901 SSTDGSFLDSVIONGVAFTRDYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960

QY 961 MYALPEKEVVYTRVDKYGTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAFLEELQAKYP 1020
Db 961 MYALPEKEVVYTRVDKYGTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAFLEELQAKYP 1020

QY 1021 ELFPARQOISTGVPMDSVKIKOWSAKYFNGTNIILGRAGVYLKQATNTYFSLVSDNTFL 1080
Db 1021 ELFPARQOISTGVPMDSVKIKOWSAKYFNGTNIILGRAGVYLKQATNTYFSLVSDNTFL 1080

QY 1081 PKSLVNPNGHGTSSVTLVFDGKGYVYVYSTSGNQAKNAFISLGNWNYFDNNGYMTGAQ 1140
Db 1081 PKSLVNPNGHGTSSVTLVFDGKGYVYVYSTSGNQAKNAFISLGNWNYFDNNGYMTGAQ 1140

QY 1141 SINGANYFSLNGIQLRNAIYDNGKVLSYGNDGRRYENGYYLFCQOQRYFQNGIMAYG 1200
Db 1141 SINGANYFSLNGIQLRNAIYDNGKVLSYGNDGRRYENGYYLFCQOQRYFQNGIMAYG 1200

QY 1201 LTRVHGAVQYFDASGFOAKGQFITTADGKLRVFDROSGNQISNRFVRNSKGWFLFDHNG 1260
Db 1201 LTRVHGAVQYFDASGFOAKGQFITTADGKLRVFDROSGNQISNRFVRNSKGWFLFDHNG 1260

QY 1261 VAVTGTVTNGQLRYFKPKNGVQAKGEFIRDANGYLRYDPPNGNEVNRNRFVRNSKGWFL 1320
Db 1261 VAVTGTVTNGQLRYFKPKNGVQAKGEFIRDANGYLRYDPPNGNEVNRNRFVRNSKGWFL 1320
```

```
QY 1321 FDHNGIANTGARVNVNGHASILSLMVFRLRESSLSQSVKVSNTMILIPKMFVIM 1375
Db 1321 FDHNGIANTGARVNVNGHASILSLMVFRLRESSLSQSVKVSNTMILIPKMFVIM 1375

RESULT 2
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4
```

```
Query Match 100.0%; Score 7143; DB 2; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MEKKVRFKLKRVKKEWTVSIAAVVTLTSLGSLVKADSTDDRQAAVTESQASLVTTS 60
Db 1 MEKKVRFKLKRVKKEWTVSIAAVVTLTSLGSLVKADSTDDRQAAVTESQASLVTTS 60

QY 61 AAKETLTATDTSTATSATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKTN 120
Db 61 AAKETLTATDTSTATSATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAKTN 120

QY 121 SDKIITTSKANRLTATGKFPANNNTAHPKVTVDKIVIPKIGKLPSSLSODDIAA 180
Db 121 SDKIITTSKANRLTATGKFPANNNTAHPKVTVDKIVIPKIGKLPSSLSODDIAA 180

QY 181 LGNVKNIIRKNGKYYKEDGTQKNYALNINNGKTFFFDETGALESNTLPSKKGNITNND 240
Db 181 LGNVKNIIRKNGKYYKEDGTQKNYALNINNGKTFFFDETGALESNTLPSKKGNITNND 240

QY 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTOKDPRPLMTWM 300
Db 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTOKDPRPLMTWM 300

QY 301 PQDETQRYVYNNNAQLGHIQTYNTATSPQLNLAQAQTIQTKIEEKITAENKTNWLRTI 360
Db 301 PQDETQRYVYNNNAQLGHIQTYNTATSPQLNLAQAQTIQTKIEEKITAENKTNWLRTI 360

QY 361 SAFVKTQSAWNSSEKPFDDHLQKGLLYSNNKLTQSAWSNYRILNRTPTNQTGKQDPR 420
Db 361 SAFVKTQSAWNSSEKPFDDHLQKGLLYSNNKLTQSAWSNYRILNRTPTNQTGKQDPR 420

QY 421 YTDARTIGGYEFLFLLANDVNSNPVQAEOLNHLFLLMNFNFIYANDPDANFOSIRVDAVD 480
Db 421 YTDARTIGGYEFLFLLANDVNSNPVQAEOLNHLFLLMNFNFIYANDPDANFOSIRVDAVD 480
```

Qy	481	NVDADLLQIAGDYLKAAKGIHKNDKAAANDHLSILLEAWSYNDTPYLHDDGDNMINMDNRLR	540
Db	481	NVDADLLQIAGDYLKAAKGIHKNDKAAANDHLSILLEAWSYNDTPYLHDDGDNMINMDNRLR	540
Qy	541	LSLLYSLAKPNQRSGMNPILTNSLVNRTDDNAETAAPVPSYSFIRAHDSVQDILIRNIIR	600
Db	541	LSLLYSLAKPNQRSGMNPILTNSLVNRTDDNAETAAPVPSYSFIRAHDSVQDILIRNIIR	600
Qy	601	TEINPNVVGYSFTTEEIKKAPETYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVG	660
Db	601	TEINPNVVGYSFTTEEIKKAPETYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVG	660
Qy	661	DMFTDDGOYMAKHTINYEAEITLLKARIKYVSGGQAMRNOQVGNSEIITSVRYGKGALKA	720
Db	661	DMFTDDGOYMAKHTINYEAEITLLKARIKYVSGGQAMRNOQVGNSEIITSVRYGKGALKA	720
Qy	721	TDGDRTRTSGVAIVIEGNPNPSBLKASDRVYVNVNGAAHKNQAYRPLLLTTDNGIKAYHS	780
Db	721	TDGDRTRTSGVAIVIEGNPNPSBLKASDRVYVNVNGAAHKNQAYRPLLLTTDNGIKAYHS	780
Qy	781	DOEAAGLVRYTNDRGELIFTAAIDIKGYANPOVSGYLGWVPVGGAAADODVRVAASTAPST	840
Db	781	DOEAAGLVRYTNDRGELIFTAAIDIKGYANPOVSGYLGWVPVGGAAADODVRVAASTAPST	840
Qy	841	DGKSVHQNAALDSRVMEFGFSNFQAFATKKEEYTNVVIKXNDKFAFASGVYTDFFEMAPQYV	900
Db	841	DGKSVHQNAALDSRVMEFGFSNFQAFATKKEEYTNVVIKXNDKFAFASGVYTDFFEMAPQYV	900
Qy	901	SSTDGSLDSVYQNGYAFTRYDYLIGISKPKNKYGTADDLVKAIKALHSKGIKWMAADWYPDQ	960
Db	901	SSTDGSLDSVYQNGYAFTRYDYLIGISKPKNKYGTADDLVKAIKALHSKGIKWMAADWYPDQ	960
Qy	961	MYALPEKEVVTATRVDKYGPVAGSQIKNTLYYVDGKSSGKDQAKYGGAFLEBLQAKYP	1020
Db	961	MYALPEKEVVTATRVDKYGPVAGSQIKNTLYYVDGKSSGKDQAKYGGAFLEBLQAKYP	1020
Qy	1021	ELFARKQISTGVPMDDPSVKIKQMSAKYPNGTNIILGRGAGYVLKQDAQNTTYFSLVSDNTFL	1080
Db	1021	ELFARKQISTGVPMDDPSVKIKQMSAKYPNGTNIILGRGAGYVLKQDAQNTTYFSLVSDNTFL	1080
Qy	1081	PKSLVNPNNHGTSSSVTGLVFDGKGYYVYTSNGQAKNAFTSLGNNWYTFDNGWYMTGAQ	1140
Db	1081	PKSLVNPNNHGTSSSVTGLVFDGKGYYVYTSNGQAKNAFTSLGNNWYTFDNGWYMTGAQ	1140
Qy	1141	SINGANYFFLSNGIQLRNAIYDNGKNVLSYDNGDRRYENGYILFGQOWYFQNGIMAVG	1200
Db	1141	SINGANYFFLSNGIQLRNAIYDNGKNVLSYDNGDRRYENGYILFGQOWYFQNGIMAVG	1200
Qy	1201	LTRVHGAVQYFDASGFOAKGQFITTADGKLYRFDRODSGNQISNRFVRNSKGFWLFDHNG	1260
Db	1201	LTRVHGAVQYFDASGFOAKGQFITTADGKLYRFDRODSGNQISNRFVRNSKGFWLFDHNG	1260
Qy	1261	VAVTGTVTENGORLYFKPNPGVQAKGEFIRDANGVLYRYDPSNGNEVNRVFRNSKGFWL	1320
Db	1261	VAVTGTVTENGORLYFKPNPGVQAKGEFIRDANGVLYRYDPSNGNEVNRVFRNSKGFWL	1320
Qy	1321	FDHNGIAVTGARVYVNGHASILSLMVFRLRESSLQSVKVGVSNTMILLPEMKFVLVM	1375
Db	1321	FDHNGIAVTGARVYVNGHASILSLMVFRLRESSLQSVKVGVSNTMILLPEMKFVLVM	1375

RESULT 3

	US-09-007-999-2	
Reason: 3	; Sequence 2, Application US/09007999	
	; Patent No. 6087559	
	; GENERAL INFORMATION:	
	; APPLICANT: Nichols, Scott E.	
	; TITLE OF INVENTION: Substitutes for Modified Starch and	
	; TITLE OF INVENTION: Latexes in Paper Manufacture	
	; FILE REFERENCE: 0356D	
	; CURRENT APPLICATION NUMBER: US/09/007,999	
	; CURRENT FILING DATE: 1998-01-16	
	; EARLIER APPLICATION NUMBER: 08/478,704	

Db 1100 DNNGYMTGAQINGVNYFSLNGLQLRDAILKNEGDTYAYYNGDRRYENGYQFMGSV 1159
Qy 1189 WRYFQNGIMAVGLTRVHGAVOYFDSAGFQAKGQFITTADGKRLYFRDRDSGNQISNRFRVN 1248
Db 1160 WRHFNNNGEMSVGLTVIDGVOYFDEMGYQAKGKEVTTADGKIRYFDFKQSGNMYNRRIEN 1219
Qy 1249 SKGEWFLFDHNGVAVTGTVTFNGQRLYFKPNKGVOAKGEFIRDANGYLYRYDPPNSGNEVRN 1308
Db 1220 BEGKWLVLGSDGAAVTSQRTINGQHLVFRANGVQVKGFEVTDHGHGRISYYDNGSGDQIRN 1279
Qy 1309 RFRVNSKGEWFLFDHNGIAVTGARVNVNGHASILSLMVF 1347
Db 1280 RFRVNAQGNFYFDNNGYAVTGARTING-----QLLYFR 1313

RESULT 5

US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 74.0%; Score 5285; DB 2; Length 1475;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;
Qy 1 MEKKVRFLKRVKRVVTVSIAASAVTLTSLGSLVKADSTDRDQAVTESQASLVTTSR 60
Db 1 MDKKVRFLKRVKRVVTVSIAASAVTLTSLGSLVKADSNESKQISNDSNTSVVTANE 60
Qy 61 AAKETLTATDTSTATSATSOPTA--TWTDNVSTTNOSTNTNTANTANFVVKPTTSBQAKT 118
Db 61 -----ESNVITEATSKQEAASQTNHTVTTSSSTSVNPKVSNPYTVGETA-- 109
Qy 119 DNSDKLIITTSKAVNRLTATGKFPVANNHTAHPKTVTDKIVPKPKIGLKKQPSLSQDDI 178
Db 110 SNGEKLQNTTVDKTSSEA-----AANNISKQTEAD-----TDVIDDSEN 149
Qy 179 AA-----LGNVKIRKNGKVVYKEDGTQLQKVALNKGKTFPFDGTGALSNTLPSKKG 234
Db 150 AANLQILEKLPNKVIEDGKYIYDNNKVRNFTLTIADGKILHFDGTGAYTDTSDITVANK 209
Qy 235 NITNNDNTNSFAQYNQVYSTDVANFEHVDHYLTAEWSYRPKYTLKDGKTWTOSTEKDPRP 294
Db 210 DIVTT--RSNLYKYNQVDRSAQSFEHVDHYLTAEWSYRPKYTLKDGKTWTOSTEKDPRP 268
Qy 295 LLMTWMPDQETQRQYVNYMNAQLGIHOTYNTATSPQLNLAAGTIQTKIEEKITAEKNTN 354
|||||

RESULT 6
US-09-008-172-2

Db 269 LLMTWMPDQETQRQYVNYMNAQLGIKNTYDDTNSQLQNLNIAAAATQAKIEAKITLLKNTD 328
Qy 355 WLQRTISAFVKTSQAMNSDSEKFPDDHLOKQGALLYSNNSKLTSQANSNYRILNRTPTNQT 414
Db 329 WLQRTISAFVKTSQAMNSDSEKFPDDHLOKQAVLYDNEGKLTPLYANSNYRILNRTPTNQT 388
Qy 415 GKXDPRTADRTTGGYEFLLANDVDSNPNVQVQAEQLNWLHFLMNFNGIYANDPDANFDSI 474
Db 389 GKXDPRTADRTTGGYEFLLANDVDSNPNVQVQAEQLNWLHFLMNFNGIYANDPDANFDSI 448
Qy 475 RVDVNVNDADLLQIAGDYLKAAKGIHKNDKAANDHLSILEANSYNDTPYLHDDGDMNIN 534
Db 449 RVDVNVNDADLLQIAGDYLKAAKGIHKNDKAANDHLSILEANSYNDTPYLHDDGDMNIN 508
Qy 535 MDNRLRLSLLYSLAKPLNORSNMPLITNSLVNRTDDNAETAAPVSYSFTRAHDSEYQDL 594
Db 509 MDNKLRLSLFLSLAKPLNQRSNMPLITNSLVNRTDDNAETAAPVSYSFTRAHDSEYQDL 568
Qy 595 IRNIIRTEINPNVVGYSFTTEEIKKAFEIYNKOLLATEKKYTHYNTALSVALLLTNKSSV 654
Db 569 IADIIKAEINPNVVGYSFTMEEIKKAFEIYNKOLLATEKKYTHYNTALSVALLLTNKSSV 628
Qy 655 PRVYVGMFTDDCGYMAHKTINYEAJETLLKARIKYVSGGQAMRNQOVGSEIITSRYG 714
Db 629 PRVYVGMFTDDCGYMAHKTINYEAJETLLKARIKYVSGGQAMRNQOVGSEIITSRYG 688
Qy 715 KGALKATDTGDRITRTSGVAVIEGNNPSLRLKASDRVVNMGAHKNQAYRPLLLTTDNG 774
Db 689 KGALKATDTGDRITRTSGVAVIEGNNPSLRLKASDRVVNMGAHKNQAYRPLLLTTDNG 748
Qy 775 IKAYHSDQEAAGLVRYTNDRGELIFTAADI KGYANPQVSGYGLVWVYVPGAAADQ--DVRV 832
Db 749 IKAYHSDQEAAGLVRYTNDRGELIFTAADI KGYANPQVSGYGLVWVYVPGAAALIKMFALRL 808
Qy 833 AASTAPSTDGKSVHQAALDSDRVMEGFSNFQAFATKBEYTNVVIKQNVDFKAEGWVTD 892
Db 809 A---REPHQQWASVHQAALDSDRVMEGFSNFQAFATKBEYTNVVIKQNVDFKAEGWVTD 865
Qy 893 FEMAPQVSVSTGSLDSDVIQNGYAFTRDYLIGISKPNKYGTADDIKVAKALHSHKGIKV 952
Db 866 FEMAPQVSVSTGSLDSDVIQNGYAFTRDYLIGISKPNKYGTADDIKVAKALHSHKGIKV 925
Qy 953 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGSKSGKQQAQYGGAF 1012
Db 926 MADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGSKSGKQQAQYGGAF 985
Qy 1013 BELOAKYPFLPARKQISTGVPMDDPSVKIKQWSAKYFNGTNILGRGAGYVLKQATNTYFS 1072
Db 986 BELOAKYPFLPARKQISTGVPMDDPSVKIKQWSAKYFNGTNILGRGAGYVLKQATNTYFN 1045
Qy 1073 LVSDN---TFLPKSLVNPNGHTSSSVTGLVFDGKGVVYVSTGNOAKNAFISLGNWYVF 1129
Db 1046 -ISDNKEINFLPKTLN-----QDSQVGSYDGKGVVYVSTSGYQAKNTFISEGDKWYVF 1099
Qy 1130 DNNGYMTGAQINGVNYFSLNGLQLRDAILKNEGDTYAYYNGDRRYENGYQFMGSV 1188
Db 1100 DNNGYMTGAQINGVNYFSLNGLQLRDAILKNEGDTYAYYNGDRRYENGYQFMGSV 1159
Qy 1189 WRYFQNGIMAVGLTRVHGAVOYFDSAGFQAKGQFITTADGKRLYFRDRDSGNQISNRFRVN 1248
Db 1160 WRHFNNNGEMSVGLTVIDGVOYFDEMGYQAKGKEVTTADGKIRYFDFKQSGNMYNRRIEN 1219
Qy 1249 SKGEWFLFDHNGIAVTGTVTFNGQRLYFKPNKGVOAKGEFIRDANGYLYRYDPPNSGNEVRN 1308
Db 1220 BEGKWLVLGSDGAAVTSQRTINGQHLVFRANGVQVKGFEVTDHGHGRISYYDNGSGDQIRN 1279
Qy 1309 RFRVNSKGEWFLFDHNGIAVTGARVNVNGHASILSLMVF 1347
Db 1280 RFRVNAQGNFYFDNNGYAVTGARTING-----QLLYFR 1313

```
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-008-172-2

Query Match          50.3%; Score 3591.5; DB 2; Length 1430;
Best Local Similarity 51.9%; Pred. No. 9.6e-254;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 MEKKVRFKLRKVRWTVTSIASAVVTL-TLSGSLVKADS-----TDDROQA 47
DB 1 METKRYKMKVKKHWTVVAVASGLITLGTTLGSSVSAETEQSDKVVTKQSEDDKA 60
QY 48 VTESQASLVTTSEA-AKETLTATDTSTATSQTATVTDNVSTTNS-----TNTTAN 101
DB 61 SESSOTDAPKTAQTEQTAQSOQANVADTSTISITKETPSNITTQANSDDKTVTNTKSE 120
QY 102 TANFVVVKPTTSEQAKTNSDKLIITTSKAVNRLTATGKVPANNNTAHPKTVTKIVPK 161
DB 121 EAQTSERTKQSEEAQTASSQALTOAKA--ELT-----152
QY 162 PKIGKLKQPSSLSQD-----DIAALGNVKNIRKNGKYYYKEDGTLQKNYALNINGKTF 216
DB 153 -----KQRTAAQAKENPNVDLAIENVK---QIDKGYIIGSDGPKPKNFALTNNKVL 203
QY 217 PFDE-TGALSNNTLPSKKNIT--NNDNTNSFAQYQNVYSTDVAFEHVDHYLTABSWTR 273
DB 204 YFDKNTGALTDTSSQYQFKQGLTKLND---YTPHQIIVNFENTSLETIDNVVTADSWR 259
QY 274 PKYILKDKTWTQSTKEDFRPLMTWPPQETQORQVNTVNAQ-LGIHOTYNTATSPLOJ 332
DB 260 PKDILKNGKWTWASSSDLRPLMSWWPKQTQIAYLNTYMOQGLGTGENYTADSSQESL 319
QY 333 NLAAQTITQIKIEKTAENKTNWLRQTI SAFVKTSAMNSDSEKPPD----DHLOKCALL 388
DB 320 NLAAQTIVQVKEIKISQTOQTQWLRIINSFVKTPQNWNSTQESDTSAGEKDHLOGALL 379
QY 389 YSNNSKLTSQANSYRILNRTPTNQTGKDPRTYADRTITGGYEFFLANDVDNSNPVQAE 448
DB 380 YSNSDK-TAYANSYRLLNRTPTSGTK--PKYFEDNSSGGYDFLLANDIDNSNPVQAE 436
QY 449 QLNWLHFLNPGNIYANDPANFDSIRVDADVNDADILQIAGDYLKAAGKHDKNDKAAN 508
DB 437 QLNWLHFLNMGYSIVANDPEANFDGVRVDADVNVNADLLQIASDYLKAHYGVDKSEKNAI 496
QY 509 DHLSTILEANSYNDTPVYLHDDGDNMIMNDNRLSLILYSLAKPLNQ-----RSGMNP 560
DB 497 NHLSTILEANSDDNPQYKNTKGQPLDINKLRLSLYALTLPLEKDAASKNEIRSGLEPV 556
QY 561 ITNSLNRNTDDNAETAAPSYSFIRAHDSVQDLIRNIIRTEINPNVVGYSFTTEIKKA 620
DB 557 ITNSLNRNRAEGKNSEMYNIFIRAHDSVQTVIAKIIKAQINPKPTDGLTFTLDELKQA 616
QY 621 FEIYNKDLATEKKYTHYNTALSYALLLTNKSQSVPRVYGDMPFTDDGQMAKTIYNAI 680
DB 617 FKYINEDMRQAKKXYTQSNIPYAYALMLSKNDSITRLYYGDMYSDDGQYMATKSPYDAI 676
QY 681 ETLLKARIKIVSGGQAMRNQOVGNSE-----LITSVRYGKGKALKATDTGDRTRTS 731
```

Qy	912	IQNGYAFPTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVDPQMYALPEKEVVT	971
Db	917	IQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSIGIQVIADWVPDQIYNLPGKEVVT	976
Qy	972	ATRVKYGTFVAGSOIKNTLYVVDGSKSGKQQAQYKGAFLBEOAKYPELFPARKQISTG	1031
Db	977	ATRVNDYGEYKDSKSEIKNTLYAAANTKSNKDYQAKYGGAFLELAAKYPSIFNRTQISNG	1036
Qy	1032	VPMDDPSVKIKQWBAKYPNGTNILGRGAGYVYLKDOATNTYFSLVSDNTFLPKSLVNPNGHT	1091
Db	1037	KKIDPEKITAWAKAFNGTNILGRGVYVYLNKNASDKYFELKGNQTYLPKQMTN-----	1091
Qy	1092	SSVTVGLVDFGKGYVYVYSTSGNOAKNAFI-SLGNMNYTFDNGYMYVTGAOSINGANYFYL	1150
Db	1092	KEASTGVNDGNGMTFYSTSGYQAKNSFVQADAKGNMYTFDNGHMYVYGLQQLNGEVYFL	1151
Qy	1151	SNGIQLRNALYDNGNKVLSYNGDGRYENGYLFF--GOOWRYFO-NGIMAVGLTRVHGA	1207
Db	1152	SNGVQLRESFLENADGSKNYFGLHGNRYSGYYSFDNDSKWRYFDASGVMAVGLKTINGN	1211
Qy	1208	VOYFDASGFOAQGFITTDGKLYRFDROSGNOISNRFRVNSKGEWFLFDHNGVAVTGTV	1267
Db	1212	TQYFDQDGGYQVKAGWITGSDGKRYFDDGSGNMAVNRFPANDKNGDWTYLNDSGIALVGVQ	1271
Qy	1268	TFGQRLYFKPQGVQAKGEFIRDANGLYRYDPNSGNEVRNFRVNSKGEWFLFDHNGIA	1327
Db	1272	TINGKTYFYQDQKQIKGIITD-NGKLKYFLANSGLARNIFATDSQNNWYFYGSDGVA	1330
Qy	1328	VTGARVNG 1336	
Db	1331	VTGSQTIAG 1339	
RESULT 8			
US-09-740-274-6			
Sequence 6, Application US/09740274			
Patent No. 6465203			
GENERAL INFORMATION:			
APPLICANT: Nichols, Scott E.			
TITLE OF INVENTION: Glucan-containing Compositions and Paper			
FILE REFERENCE: 0357CRD			
CURRENT APPLICATION NUMBER: US/09/740,274			
CURRENT FILING DATE: 2000-12-19			
PRIOR APPLICATION NUMBER: 09/210,361			
PRIOR FILING DATE: 1998-12-11			
PRIOR APPLICATION NUMBER: 09/007,999			
PRIOR FILING DATE: 1998-01-16			
PRIOR APPLICATION NUMBER: 08/478,704			
PRIOR FILING DATE: 1995-06-07			
PRIOR APPLICATION NUMBER: 09/009,620			
PRIOR FILING DATE: 1998-01-20			
PRIOR APPLICATION NUMBER: 08/485,243			
PRIOR FILING DATE: 1995-06-07			
PRIOR APPLICATION NUMBER: 09/008,172			
PRIOR FILING DATE: 1998-01-16			
PRIOR APPLICATION NUMBER: 08/482,711			
PRIOR FILING DATE: 1995-06-07			
NUMBER OF SEQ ID NOS: 6			
SOFTWARE: FastSeq for Windows Version 3.0			
SEQ ID NO 6			
LENGTH: 1430			
TYPE: PRT			
ORGANISM: streptococcus mutans			
US-09-740-274-6			
Query Match			
Best Local Similarity 50.3%; Score 3591.5; DB 2; Length 1430;			
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;			
Qy	1	MEKKVRFKLKRVKRWTVTSIASAVVTL-TLSGSLVKADS-----TDDRQQA	47
Db	1	METKRYKMKHKHWTVAASGLITLGTTLGSSVSAETEQTSDKVTQKSEDDKAA	60
Qy	48	VTSQASLVTTSEA-AKETLTATDTSTATSATOPTAIVTDNVSTTQNS-----TNTAN	101
Db	61	SESSQTDAPKTKQAQTEQQAQSANVADTSTSTIKETPSQNIITQANSDDKVTNTKSE	120
Qy	102	TANFVVKPTTTSQAOKTDSNDKIITTSKANRLTATGKFVPANNNTAHPKTVTDKIVPIK	161
Db	121	EATQSEBRTKQSEAOQTASSQALTOAKA--ELT-----	152
Qy	162	PKTGKLUKQPSLSQD-----DIAALGNVKNIRKNGKYVYKEDGTLOKNYALNINNGKTF	216
Db	153	-----KQRTAAQAKENKPNVDLAAIPNVK--QIDGKYVYIGSDGQPKKNFALTNNKVL	203
Qy	217	FPDE-TGALSNNLTLPKKGKGNIT--NNDNTNSFAQYQVYSTDVANFEHVDHYLTAESWYR	273
Db	204	YFDKNTGALTDTSQYQFKQGLTKLND---YTPHNQIVNFENTSLETIDNYVTADSWYR	259
Qy	274	PKYILKDGKTKWTOSTEKDPFLLMTWPDQETORQVYVYNNAAQ-LGIHQTYNTATSPQLQ	332
Db	260	PKDILKNGKTKWTASSEDLRPLLMSWMPDKQTQIAYLNNYNNQQGLGTGENYTTADSQESL	319
Qy	333	NLAAQTIQTKIEBKITAENKTNMLROTISAFVKTQSAWNSDSKPPD----DHLQKALL	388
Db	320	NLAAQTVQVKIEKISQTOQTQWLRIINSFVKTPQWNSQTESDTSAGEKHQLOGCALL	379
Qy	389	YSNNSKLTQANSRYILNRTPTNQTKGKDPRTADRTIGGYEFLLANDVDSNPVVQAE	448
Db	380	YSNSDK-TAYANSKYLLNRTPTSQTK--PKYFEDNSSGGYDFLLANDIDNSNPVVQAE	436
Qy	449	QLNLWHLFNLFGNIYANDPDANFDSIRVDADVNDADLQIAGDYILKAAKGHIHNDKAA	508
Db	437	QLNLWHLNMYGSIIVANDPEANFDGVRVDADVNDVNDLQIAGDYILKAHYGVDSKNAI	496
Qy	509	DHLSILEANSYNDTPYLLHDDGDNNIMNRLRLSLYSLAKPLNQ-----RSGMNP	560
Db	497	NHLSILEANSNDPQYNKDTKGAQLPDNKLRLSLYALTLPLEKDAKNNEIRSGLEPV	556
Qy	561	ITNSLVNRTDDNAETAAPVSYSPIRAHDSQVQDLIRNIIRTEINPNVVGYSFTTEIRIKA	620
Db	557	ITNSLVNRSAGKNSERMANYIFIRAHDSQVQTVIAKIIKAQINPKTDGLTFLDELQA	616
Qy	621	FEIYNKDLATEKKYTHYNTALSYALLTNKSSVPVYGDMPDQGMMAHKTINYEAI	680
Db	617	FKIYNEDMRQAKKYYTQSNIPATAYALMNLKSDISITRLYGDYSDGQYMATKSPYDAI	676
Qy	681	ETLLKARIKYVSGGAWRNQVCNSE-----IITSVRCKGALKATDGTDRTRTS	731
Db	677	DTLLKARIYAGGQDMKITTYVEGDKSHMDWDYTVLTTSVRGTGANEATDQGEATKQ	736
Qy	732	GVAVIEGNPNPSLRLKASDRVNVNMGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRYT	791
Db	737	GVAVITSNPNPSLKLNDKVINVMGAAHKNQAYRPLLLTTKGLTSYTSDAAKSLYRKT	796
Qy	792	NDRGELIFTAADIKGYANPOVSGYLGVWVPVGAADQDVRVAASTPSTDGKSVHQAAL	851
Db	797	NDKGELVFPASDIQGYLNFQVSGYLAVWVPVGAASDNDQVRVAASNKANATGVYESSAL	856
Qy	852	DSRVMEGESNFOAFATKEEYTNVVIKAVNDKFAEWGVTDEMAPPQYVSSDTGSFLDSV	911
Db	857	DSQLIEGFSNFDQFVTKDSYTNKKIAQNVQLFKSWGVTSEMAPQYVSSDSDGSLDSI	916

```
QY 48 VTESQASLTTSEA-AKETLTATDTSTATSATSQPTATVTDNVSTTNSQ-----TNTTAN 101
Db 61 SSSQTDAPKTKQAQTEQTOAQSQANVADTSTSTIKETPSQNIITQANSDDKTVTNTKSE 120
QY 102 TANFVVKPTTTSQAQKTSNDKIIITTSKAVNRLTATGKFPANNTAHPKTVTDKIVPIK 161
Db 121 EACTSERTKQSEAEATTASSQALQAKA--ELT----- 152
QY 162 PKIGKLKQSSLSQD-----DIAALGNVKNIRKVNKGYKYYKEDGFLQRYALNINGKTF 216
Db 153 -----KORQTAQAENKPNVDLAAIPNVK---QIDGKYIYISGSGQPKKNFALTNNKYL 203
QY 217 FPDE-TGALSNNLTPSKKNIT--NNDNTNSPAQYNQVYSTDVANFEHVDHYLTAEWSYR 273
Db 204 YFDKNTGALTDSYQYFKQGLTKLND---YTPHNQIVNFENTSLKETIDNVYVADSWYR 259
QY 274 PKYILKDGKTKTOSTEKDFRPLMTWMPDQETQRYVYVYMAQ-LGIHOTYNTATSPLOL 332
Db 260 PKDILKNGKWTWASSSDRLPILMSWPKQTOIAVLTWNOQGLGTGENYVADSSQESL 319
QY 333 NLAAQTIQTKIEKI TAEKNTNWLRTISA FVKTSQAMNSDSEKPPD-----DHLQKCALL 388
Db 320 NLAAQTQVQKIETKISQTOQTQWLRLDIINSFVKTPQPNWNSQTESDTSAGEKDHLQGCALL 379
QY 389 YSNSKLTQSAQSNRYRILNRTPTNQTGKDPRTYADRTIGGYEFLANDVDSNPVQAE 448
Db 380 YSNSDK-TAYANSDYRLNRTPTTSQTK--PKYFEDNSGGYDFLLANDIDNSNPVQAE 436
QY 449 QLNWLHFLNFGNIYANDPDANFDSIRVDADVNDADLLQIAGDYLLKAAKGHIKNDKAAN 508
Db 437 QLNWLHLYMNGSIVANDPEANFDPGRVDADVNDVNDADLLQIASDYLLKAHGVQDKSEKAI 496
QY 509 DHLSTLEAWSYNDTPYLDHGDGNMNMNRLRLSLYSLAKPLNQ-----RSGMNPFL 560
Db 497 NHLSTLEAWSNDPQYNKDTKGAQLPDKNKLRLSLYALTLPLEKDNKNEIRSGLEPV 556
QY 561 ITNSLVNRTDDNAETAAPVSYSFIRAHDSQVQDLINRIIRTEINPNVVGYSFTTEIKKA 620
Db 557 ITNSLNRSAEGKNSERMANIYFIRAHDSQVQVIAKIIKAQINPKTDLTPTFLDELKQA 616
QY 621 FEYNKDLATEKKTHTYNTALSYALLTNKSSVPRVYVYGDMDFTDQGYMAHKTINYEAI 680
Db 617 FKLYNDEMOKAKKYTQSNIPYALMLNKSQITRLYGDWYSDQGYWATKSPYIDAI 676
QY 681 ETLLKRIKYVSGQAWRNQQVGNSE-----IITSVRYGKGALKATDTGDRTRTS 731
Db 677 DTLLKARIKYAAGQDMKITVYEGDKSHMDWDVTGVLTSVRYGTGANEATDQSEATKQ 736
QY 732 GYAVIEGNPNPSLELKASDRVYVNMGAHKNQAYRPLLLTTDNGIKAYHSDQEAAGLVRVT 791
Db 737 GMAVITSNPNPSLKLNQNDKVIYVNMGAHKNQOYRPLLLTTKQGLTSTYSDAAAKSLYRKT 796
QY 792 NDRGELIFTAADIKGYPANQVSGYLVVYVPGAAADQVRAASTAPSTDGKSVHONAAAL 851
Db 797 NDKGELVFDASDIQGLYINQVSGILAVVYVPGASNDQDVRVAASNKANATQGYVESSSAL 856
QY 852 DSRVMEFGFSNFQAFATKEEYTNVVIKNDVKFAEWGVTDFEMAPQYVVSSTDGSFLDSV 911
Db 857 DSQLIYEGFSNFQDFVTKSDYTNKKIAQNVQLFKSWGVTSPFMAQYVVSSEDSFLDSI 916
QY 912 IQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHKSIGIKWADWVPDQMYALPEKEVVT 971
Db 917 IQNGYAFEDRYDLAMS KNNKYSQQDMINAVKALHKSIGIQTADWYPDQIYNLPKGEVVT 976
QY 972 ATRVDKYGTPVAGSQIKNTLYVVDGKSGKQDQAKYGGAFLELOAKYPELPARKOISTG 1031
Db 977 ATRVDYGEYRKDSEIKNTLYAANTKSNKQDYQAKYGGAFLELAKEYSIFNRTQISNG 1036
QY 1032 VPMDPSVKIKOMSAKYFNGTNIILGRGAGYVLYKQDQATNTYFSLVSDNTFLPKSLVNPNHGT 1091
Db 1037 KKIDPSEKITA WAKAYFNGTNIILGRGVYVLYKDNASDKYFELKGNQTYLPKQMTN----- 1091
QY 1092 SSSVTGLVFDGKGYVYVYSGNQAKNAFI-SLGNWYVYFDNNGYMWYTGASINGANYFYFL 1150
```

```
Db 1092 KEASTGFVNDGNGMTFYSTSGYQAKNSFVDQDAKGHWYFDDNGHWYGLQQLNGEVOYFL 1151
QY 1151 SNGIOLRNAIYDNGKNVLSYNGNDGRRYENGYYLF--GQOMRYFQ-NGIMAYGLTRVHGA 1207
Db 1152 SNGVQLRESFLENADGSKNYFCHLGNRYSNGYYSFDNDSKWRYPDASGVMAVGLKTINGN 1211
QY 1208 VOYPDASFOAKQOIFITTADGKLRYFDRDSGNQISNRFVRNSKGEWFLFDHNGVATGTV 1267
Db 1212 TOYFDQDGYQVKGAWITGSDGKKRYFDDSGNMAVNFANDXNGDWYLLNSDGIALLVGVO 1271
QY 1268 TENGORLYFKPNGVOAKGEFIRDANGLYRYDYPNSGNEVRNRFVRNSKGEWFLFDHNGIA 1327
Db 1272 TINGKTYFGQDGKQIKGKIITD-NGKLIFLANGELARNIPATDSQNNWYFFGSDGVA 1330
QY 1328 VTGARVAVNG 1336
Db 1331 VTGQTIAG 1339
```

RESULT 9

```
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
; STREET: Level 8, 168 Walker Street
; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,824
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1577 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
;
US-08-793-824-2
```

```
Query Match 45.0%; Score 3216; DB 1; Length 1577;
Best Local Similarity 47.7%; Pred. No. 3.6e-226;
Matches 687; Conservative 214; Mismatches 413; Indels 126; Gaps 33;

QY 1 MEKKVRFKLRKVKGRWTVTSIASAVVVTLSLGS-----SLVKADST-----DDROQAVT 49
Db 1 MENKVRFKLRKVKKNWTVIGVTT--LSMVALAGGSLAQGKVEADETSPNGDGLQQLSE 58
```



```
Qy 528 DGDNMNDRRLSLLSYLAKPLNORSQ-----NPLITSLAV 566
Db 686 QGNAQLTWDTYVTSFGNSLTHGANNRSMWFLDTGYLNGDLNKKIVDKRPNMSGTLV 745
Qy 567 NRTDNDAAATAAPVSFIRAHSEVDQILIRNIIRTEINPNVVG---SFTTEEIKKAFEI 623
Db 746 NRANSQGTVPINPSFVRADHYDAQDPIR---KAWIDHGIKNQMDFTFPOLAQGMF 802
Qy 624 YNKOLLATE--KYYTHYNALSYALLTNKSSVPRVYGDMPDTDGOYMAHKTINYEAI 681
Db 803 YKQDQENPSGFKYNDYNLPSAYAMLLTNKOTVPRVYGDMPYLEGQYMEKGTIYNPVS 862
Qy 682 TLLKARIKVVSGQAM-----RNOQVGNSEIITSRYCKGALKATDGDRTT----- 728
Db 863 ALLKARIKVVSGQMATDSSGKDLKDGEDLLTSVRFKGIM-----TSDQTTTODNSQD 918
Qy 729 -RTSGVAVIEGNPNRLKASDRVVVNMGAHKNQAYRPLLLITDNGIKAYHSDQEAAGL 787
Db 919 YKQIGIVGVNPNPLKLNNDKTIILHMGAKHKNQLYRALVLSNDSGIDVYDSDDKAPT 978
Qy 788 VRYTNDRELIPATA-----DIKYANPQVSGYLGWVWPVGAADQVR-VA 833
Db 979 --RTNDNGDLIPHKNTFVKQDGTIINVMKGSNALLISGYLGWVWPVVGASDQARTVA 1036
Qy 834 ASTAPSTGKSVHQAALDSRVMPGFGSNFQAFATKKEYNVVIKAKNDKPAEWGVTDF 893
Db 1037 TESSSSNDGVSFHSNAALDSNVYRFGSNFQAMPTSPSQSTNVVIATKANLPKEIGITSF 1096
Qy 894 ENAPOYVSTDG-----SPLDSVIQNGYAFTRDYDLGISK-----PNKYGTADDLVRAI 942
Db 1097 ELAPOYRSGDNYGMSFLDSFLANGYAFTRDYDLGFNKADGNPNPKYGTDDQLRNAI 1156
Qy 943 KALHSGIKGMADWPDQMYALPEKEVVTATRVKYGTFPVAGSQKNTLYVVDGKSSGD 1002
Db 1157 EALHKNQMAIADWPDQIYALPGKEVVTATRVDSRGNQLKDTDFVNLLYVANTKSSGD 1216
Qy 1003 QOAKYGGAFLBLOAKYFELPARKQISTCVPMDPSVKIKOMSAKYFNGTNTILGRGAYVL 1062
Db 1217 YOAKYGGEBFLDKREYPSLKFQNGVSTGQPIDASTKIKOMSAKYMNGTNTILHRGAYVL 1276
Qy 1063 QDAQNTYFSLVSDN--TLPKSLVNPNGHTSSVTGLVFDGKGYVYVYSGNQAKNAFIS 1121
Db 1277 KDAWNTQVFNIAKTNVEFLPQLQN-----KDAQTGFISSAGVYKYSISGVQAKDTIE 1331
Qy 1122 LGN-WNYFDNNGYMTGAQ-----SIN--GANYFYSNGIQLR-----NA 1159
Db 1332 DGNGNWYFDKGYMVRSGQGENPIRTVETSVNTRNGNYFMPNGVELRKGFPTDNGSNV 1391
Qy 1160 IY--DNGKNVLSYNGDGERYNGYVLFQOQWRYFQNGIMAVGLTRVHG-AVOYFDASGF 1216
Db 1392 YFDDQKKVRDKYINDDA--NNFYHL-----NVDGTMRSGLFKFSDTLQYFASNGV 1442
Qy 1217 QAKGQFIATTADGLRYFDRDSQNOISNRFVRNSKGEWFLFDHNGVAVT-----GTVTFN 1270
Db 1443 QIKDSYAKDSKGNKYFYSATGN-----NDTGAQTWDGNGYIITSDANNITGVN 1494
Qy 1271 GORLYFKPNGVOAKGEF-----IRDANG-----YLRYYDPNSGNEVR-NRFRNSK 1315
Db 1495 TDYATITSSLRDEGLFANAPYGVVTVKQNGNDLKQWYINHTKYEGQGVQVTRQYTD 1554
Qy 1316 G-EWFLFDHNGIYAVTGARV 1333
Db 1555 GVSMNLITFAGDGLQGORL 1573
```

RESULT 11

US-09-604-957-3

; Sequence 3, Application US/09604957

; Patent No. 6486314

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAUI, HAKIM

```
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3
```

Query Match 32.9%; Score 2347.5; DB 2; Length 1278;

Best Local Similarity 41.3%; Pred. No. 1.1e-162;

Matches 538; Conservative 181; Mismatches 408; Indels 175; Gaps 35;

```
Qy 29 TSLSGSLVKADSTDDRQAAVTESQASLVTTSEAAKETLTATDTSTATSQPTATVTDN 88
Db 46 SAVSGFDFTIKLTNDYQALNGQLQVLLRFSKAADG--NPSGDNVTVDQFSKNYATG 103
Qy 89 VS-----TTNQS-----TNTTANTA-----NFFV 107
Db 104 FDYKVNQGVFSGWHATNQSNDKDSQWIIIVLVNGKEVKQLVNDTKEGAAGFRNDVY 163
Qy 108 KPPTTSEQAKTNSDKIITTSKAVNR-----LTATGKFVPANNNTAHPKTVT 154
Db 164 KVNPAIENSSMSGFGIITLPVTVKNEVQLVHRFSNDVKTEGNYV-----DFW 213
Qy 155 DKYIPKPKI---GKLQPSLSQDDTAALGNVKNIRKVNKYYV--KEDGTLKNYAL 209
Db 214 SELMPVKDSFOKNGPLKQ-----FGLQTINGQQYYIDPTTGQPRKNELL 258
Qy 210 NINGKTFEFTGALSNNTLPSK--KGNITNNDNTNSFAQVNOVSTDVANPEHVDHYLT 267
Db 259 QSGNNWIYFSDTGVGTWALBLOFAKGTVSNE---QVRNGNAAYSDDKSIENVNGILT 315
Qy 268 ABSWYRPKYILKDGKWTQSTKDFRPLMTWMPDQETORQOVYVYMNAAQLGHIQT---Y 323
Db 316 ADTWYRPQKILKDGTTTDSKETDMRPILMWVWPNLTLOAYVLYNTMKQGNLLPSALPF 375
Qy 324 NTATSPLOLNLAQTIQKIEKITAENKTNWLRQTI SAFVKTSQANNSDSEKPDHLO 383
Db 376 NADADPAELNHYSEIVQQNIEKRISETGNTDMLRTEHDFVTNPNMWNKDSNVNFSGIQ 435
Qy 384 -KGALLYNNKSLTSQANSNRYILNRTPNTQTKKPRVTADRTIGGYEFFLIANDVDSN 442
Db 436 FQGGFLKYENSGLTPYANSDFRLLGRMPIN-----IKDQYRGQEFLLANDIDSN 486
Qy 443 PVVQAEQLNWLHFLMNFNIYANDPDANFDSIRVDADVNDADLQIAGDYLLAKAKGIHK 502
Db 487 PVVQAEQLNWLVLNLFNGTITANNDQANFDSVRVDAPDNIDADLMNIAQDYENAAVGM-D 545
Qy 503 NDKAANDHLSILEAWSYNDTPLYHDDGDNMNMNRRLSLLSYLAKPLNQBSGNPLIT 562
Db 546 SDAVSNKHINLEDMNHADPEYFNKIGNPQLTMDPTIKNSLNHGLSDATN-RWGLDAIVH 604
Qy 563 NSLVNRTDDNABTAAPSPYSFIRAHDSVQDLIRNIIRTEINPNVVG---YSFTTEEIKK 619
Db 605 QSLADRENNSNTEVVI PNYSFVRAHDNNSQIQINAIR-----DVTGKDYHTFTFEDEQ 659
Qy 620 AFEIYNKOLLATEKRYTHYNTALSYALLTNKSSVPRVYGDMPDTDGOYMAHKTINYE 679
Db 660 GIDAYIDQNSNTVKYKYNLYNPASVAILLTNKTDTIPRVYGDLYTDGGYMEHQTRYDT 719
Qy 680 IETLLKARIKYVSGQAMRNOQV--CNSEIIITSVRKYGKALKATDGTTRTSGVAVIEG 738
Db 720 LTNLLKSRVKYVAGQSGMQTMSVGGNNNLTSVRYGKGMTATDGTDETRYGGIGVVVS 779
Qy 739 NNPSRLKASDRVVVNMGAHKNQAYRPLLLITDNGIKAYHSDQAAAGLVRYTNDRGEIL 798
```


Db 780 NTPNLKLVNDKVVLLHMGAAHKNQYRAAVALTTTDDGVINTSDQAP--VAMTDENGDLV 837
QY 799 FTA-----AD--IKGYANPOVSGYLGWVVPVGAADQDVRVAASAPSTDGKSVH 846
Db 838 LSSHNLVWNGKEADTAVQYANPDVSGYLVAVVPGASDNQDARTAPSTKSGNSAYR 897
QY 847 QNAALDSRVMPGFSNFOAPATKKEEYTNVVIKAVNDKFAEWGVTDFEMAPQVVSSTDS 906
Db 898 TNAAFDSNVIFEAFSNFVYPTKESERANVRIAQNAADFFASLGFTSFEMAPQVNSKORT 957
QY 907 FLDSVQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVDPQYALPE 966
Db 958 FLDSVQNGYAFTRDYDLGISENENKYGTDEDLNAIQALHKGAGLVMAWVDPQYALPE 1017
QY 967 KEVVTATRVKDYGTVPVAGSQIKNTLYVVDGKSGKQQAQKYGGAFLLEELQAKYPELFARK 1026
Db 1018 KEVATVTRVDRGNWVKDAIINNLYVNVN-TIGGGEYQKYGGAFLDKLQKLYPEIFTK 1076
QY 1027 QISTGVPMDPSVKIKOWSAKYFNGTILGRGAGYVLKQATNTYFSLVSDNTFLPKSLV- 1085
Db 1077 QVSTGVAIDPSQKITEWSAKYFNGTILHRSGYVLKADGGQYVNLGTTTKOFLPQLTG 1136
QY 1086 NPNHGTSSSVTGLVFDGKGYVYVYSTSGNOAKNAFI--SLGNWYVFDNNGYVMTGAQSN 1143
Db 1137 EKKQNEGVK--NDG-NYFYDLAGNWKVTFIEDSVG-NWYFDDQKGVENKHFVD 1192
QY 1144 -----GANYFSLNGIOLRNAIYDNGNKVLSYNGDGRYENGYYLFGQWRYFQNG- 1196
Db 1193 VDSYGEKGYFFLKGVSFRGLVQT-----DNGTY-----YFDNYCK 1230
QY 1197 MAVGLTRVHGAVQY-FDASG-----FOAKGQFITTAD-GKL 1230
Db 1231 MYRNOTINAGAMIYTLDENGKLIKASYNDAEYPTSTDVGKM 1272

RESULT 12
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRN
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match 32.9%; Score 2347.5; DB 2; Length 1781;
Best Local Similarity 41.3%; Pred. No. 1.9e-162;
Matches 538; Conservative 181; Mismatches 408; Indels 175; Gaps 35;

QY 29 TSLSSGLVKAUSTDRQOAVTESQASLVTTSEAAKETLTATDTSTATSQPTATVTDN 88
Db 549 SAVSGFDTTIKLTNDQYQALNGQLVLLRFSKAADG--NPSGDNTVTDQFSKNYATTG 606
QY 89 VS-----TTNOS-----TWNTANTA-----NFVV 107
Db 607 FDYKVNQNGVEFSGHATNSNDKDSQWIVLVNGKEVKQLVNDTKEGAGFNRNDVY 666
QY 108 KPTTSEQATDNSDKIITTSKAVNR-----LTATGKFPVANNNTAHPKTVT 154

Db 667 KVNPAIENSSMGFOGIITLPTVTVKXNENQVLRHFRFSNDVKTGEGNVY-----DFW 716
QY 155 DKIVPIKPKI---GKLKOPSSLSQDDIAALGNVKNIRKNGKYYY-KEDGTLOKNYAL 209
Db 717 SELMPVVDKDFQKNGPLKQ-----FGLQTINGQOYIDPTTGPGRKNFL 761
QY 210 NINGKTFEFDGTCALSNNTLPSK--KGNITNNDNTNSFAQYNQVYSTDVANFHVHLYT 267
Db 762 QSGNWIYFSDTGVGTNALELQFAKGTVSSNE---QYRNGNAAYSDDKSIENVNGYLT 818
QY 268 AESWTRPKVILKDGKTWTOSTEKDPRPLMTWPDQOETORQYVYNNAQLGIIHOT---Y 323
Db 819 ADWTRPKOILKDGTTTDSKETDMRPIILMVWMPNTLTQAYLYNMKQHGNLPSALPFF 878
QY 324 NTATSPQLNLAAOTIQTKEIKTAEKTNMLRQTIISAFVKTQSAWNSDSEKPPDDHLQ 383
Db 879 NADADPAELNHYSEIVQOQNIKRISETGNTMLTLMHDFVTNNPMNKNKDNVNFSGIQ 938
QY 384 -KGALLYSNNSKLTSQANSYRILARTPTNQTGKDPRTADRTIGGYEFLANDVNSN 442
Db 939 FQGGFLKYENSDLTPVANSDYRLGRMPIN-----IKOQYRGQEFLLANDIDNSN 989
QY 443 PVQOAEQLNHLPLNFGNIYANDPANFDSIRVDVNDVADLLQIAGDYLKAAGIHK 502
Db 990 PVQOAEQLNLYLLNFGTITANNQANFDSVRVDAPDNIDADLNMNIAQDYFNAAVGM-D 1048
QY 503 NDKAANDHLSILAEASYNDDTPYLHDDGDNWIMNDNRLSLLSYLAKPLNQRSGMPLIT 562
Db 1049 SDVSNKHINILEDNWHADPEYFNKIGNPQLTMDDTIKNSLNHLSDATN-RWGLDAIVH 1107
QY 563 NSLVNRTDDNAETAAPVPSYFIRAHDSQVQDLIRNIIRTEINPNVVG---YSFTTEIKK 619
Db 1108 QSLADRENSTENVIPNYSFVRAHDNNSQOIQNAIR-----DVTGKYHYHFTFDEBQK 1162
QY 620 AFEIYNKOLLATEKYYTHYNTALSYALLTNKSSVPRVYVYGMFTDDGQYMAHKTINYE 679
Db 1163 GIDAYIQONSIVKKNLYNIPASYAILLTNKTDTIPRVYVYDLYTDGQYMEHOTRYDT 1222
QY 680 IETLLKARIKYVSGGQAMRNQOV-GNSEIITSVRYGKALKATDGTDRTRTSQVAVIEG 738
Db 1223 LTNLLKSRVYVAGGQSMQMTMSVGGNNILTSVRYGKAMTATDGTDETRTQIGVYVS 1282
QY 739 NNPSLRLKASDRVVVNMGAHKNQYRPLLLTTDNGIKAYHSDQEAAGLVRTVNDRGELI 798
Db 1293 NTPNLKLVNDKVVLLHMGAAHKNQYRAAVALTTTDDGVINTSDQAP--VAMTDENGDLV 1340
QY 799 FTA-----AD--IKGYANPOVSGYLGWVVPVGAADQDVRVAASAPSTDGKSVH 846
Db 1341 LSSHNLVWNGKEADTAVQYANPDVSGYLVAVVPGASDNQDARTAPSTKSGNSAYR 1400
QY 847 QNAALDSRVMPGFSNFOAPATKKEEYTNVVIKAVNDKFAEWGVTDFEMAPQVVSSTDS 906
Db 1401 TNAAFDSNVIFEAFSNFVYPTKESERANVRIAQNAADFFASLGFTSFEMAPQVNSKORT 1460
QY 907 FLDSVQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVDPQYALPE 966
Db 1461 FLDSVQNGYAFTRDYDLGISENENKYGTDEDLNAIQALHKGAGLVMAWVDPQYALPE 1520
QY 967 KEVVTATRVKDYGTVPVAGSQIKNTLYVVDGKSGKQQAQKYGGAFLLEELQAKYPELFARK 1026
Db 1521 KEVATVTRVDRGNWVKDAIINNLYVNVN-TIGGGEYQKYGGAFLDKLQKLYPEIFTK 1579
QY 1027 QISTGVPMDPSVKIKOWSAKYFNGTILGRGAGYVLKQATNTYFSLVSDNTFLPKSLV- 1085
Db 1580 QVSTGVAIDPSQKITEWSAKYFNGTILHRSGYVLKADGGQYVNLGTTTKOFLPQLTG 1639
QY 1086 NPNHGTSSSVTGLVFDGKGYVYVYSTSGNOAKNAFI--SLGNWYVFDNNGYVMTGAQSN 1143
Db 1640 EKKQNEGVK--NDG-NYFYDLAGNWKVTFIEDSVG-NWYFDDQKGVENKHFVD 1695
QY 1144 -----GANYFSLNGIOLRNAIYDNGNKVLSYNGDGRYENGYYLFGQWRYFQNG- 1196

TYPE: PRT
ORGANISM: Streptococcus mutans
US-09-995-749A-10

Query Match 23.7%; Score 1690.5; DB 2; Length 545;
Best Local Similarity 60.1%; Pred. No. 3.7e-115;
Matches 328; Conservative 84; Mismatches 115; Indels 19; Gaps 4;

QY 433 LLANDVNSNPVQAQQLNLHFLMNFNGIYANDPDANFDSIRVDVNDVADLLQIAGD 492
Db 1 LLANDIDNSNPVQAQQLNLHFLMNYGSIANDPEANFDGVRVDVNDVADLLQIAGD 60

QY 493 YLKAAGIHKNDKANDHLSILEASYNTPYLLHDDGDMNNDNRRLSLYSYLAKPLN 552
Db 61 YLKAHGVDKSEKNAINHLSILEAWSNDPQYNKOTKGAQLPIDNKLRLSLYALTRPLE 120

QY 553 Q-----RSGMNPILTNLSVNRDNDNAETAAPSVSFIRAHDSVQDLIRNIIRTEIN 604
Db 121 KQASKNEIRSGLEPVITNSLNRSAEGKNSEKMANIIFIRAHDSVQTVIAKIIKAQIN 180

QY 605 PNWVGYSFTTEIKKAFIYNKDLALATEKKYTHYNTALSYALLTNKSSVPRVYGDMEF 664
Db 181 PKTDGLTFLDELKQAFKIYNEDMRQAKKKYQSNIPYALMLSNKDSITRLYIGDMS 240

QY 665 DGGQYMAHTINYEAETLLKARIKIVSGQAMRNQGVNSE-----IITSVRYGK 715
Db 241 DGGQYMATKSPYYDAIDTLLKARIKIVAGGQDMKITYVEGDKSHMDWDVTGLTSVRYGT 300

QY 716 GALKATDGTGRTTRISGVAVIEGNNPSLRKASDRVVMGAAHKNQAYRPLLLTTDNGI 775
Db 301 GANEATDQGEATKQGMVITSNPSLKLNDKQVIVNMGAAHKNQYRPLLLTTKQGL 360

QY 776 KAYHSDQEAAGLVRYTNRGELIFTAADIKG-YANPQVSGYLGVVVPVGAADQDVRAA 834
Db 361 TSYTSDAAKSLYRKNDKGLVFDASDIQGLYLPQVSG-LAVWVPVGSNDQDVRAA 419

QY 835 STAPSTDGKSVHONAAALDSRVNPEGSFNFQAFATKKEEYTNVVIKQNDKFAEWGVTDFE 894
Db 420 SNKANATGVYESSGALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVOLFKSGWTSFE 479

QY 895 MAPQYVSSDGSFLDSVIONGYAFTDRYDLGISKNKYGTADDLVKAKALHSGIKVMA 954
Db 480 MAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHSGIQVIA 539

QY 955 DWVPDQ 960
Db 540 DWVPDQ 545

Search completed: February 11, 2006, 19:45:39
Job time : 35.1047 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 123.071 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-35
Perfect score: 7143
Sequence: 1 MEKKVRFKLKRVKRWTVTS.....VKVSNVTMLIPMKFVIM 1375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7143	100.0	1375	3	US-09-740-274-4
2	7143	100.0	1375	4	US-10-383-930-35
3	7143	100.0	1375	5	US-10-797-821-35
4	5285	74.0	1475	3	US-09-740-274-2
5	5285	74.0	1475	4	US-10-383-930-34
6	5285	74.0	1475	5	US-10-797-821-34
7	4214.5	59.0	1590	4	US-10-383-930-37
8	4214.5	59.0	1590	5	US-10-797-821-37
9	3591.5	50.3	1430	3	US-09-740-274-6
10	3591.5	50.3	1430	4	US-10-383-930-36
11	3591.5	50.3	1430	5	US-10-797-821-36
12	3161.5	44.3	1554	4	US-10-383-930-38
13	3161.5	44.3	1554	5	US-10-797-821-38
14	3113	43.6	1497	5	US-10-484-218-18
15	3070	43.0	1518	4	US-10-383-930-40
16	2893.5	40.5	1365	4	US-10-797-821-40
17	2893.5	40.5	1365	5	US-10-383-930-39
18	2893.5	40.5	1365	5	US-10-797-821-39
19	2861	40.1	1595	5	US-10-484-218-20
20	2501.5	35.0	2057	4	US-10-417-280A-2
21	2410	33.7	1006	5	US-10-484-218-22
22	2383.5	33.4	1771	5	US-10-484-218-14
23	2383.5	33.4	1771	5	US-10-484-218-12
24	2347.5	32.9	1781	3	US-09-995-749A-2
25	1690.5	23.7	545	3	US-09-995-749A-10
26	1616	22.6	522	3	US-09-995-749A-11
27	1445.5	20.2	535	3	US-09-995-749A-13

28	1421	19.9	787	5	US-10-484-218-16	Sequence 16, Appl
29	1337	18.7	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1273.5	17.8	525	5	US-10-484-218-23	Sequence 23, Appl
31	677.5	9.5	224	5	US-10-484-218-4	Sequence 4, Appl
32	673	9.4	223	5	US-10-484-218-6	Sequence 6, Appl
33	641	9.0	223	5	US-10-484-218-10	Sequence 10, Appl
34	585	8.2	221	5	US-10-484-218-2	Sequence 2, Appl
35	553	7.7	221	5	US-10-484-218-8	Sequence 8, Appl
36	312.5	4.4	2710	4	US-10-011-366-6	Sequence 6, Appl
37	312.5	4.4	2710	4	US-10-354-774-6	Sequence 6, Appl
38	312.5	4.4	2710	4	US-10-271-013-6	Sequence 6, Appl
39	312.5	4.4	2710	4	US-10-729-122-6	Sequence 6, Appl
40	312.5	4.4	2710	4	US-10-729-039-6	Sequence 6, Appl
41	312.5	4.4	2710	5	US-10-729-527-6	Sequence 6, Appl
42	312.5	4.4	2710	5	US-10-727-898-6	Sequence 6, Appl
43	312.5	4.4	2710	5	US-10-728-696-6	Sequence 6, Appl
44	312.5	4.4	2710	6	US-11-001-241-6	Sequence 6, Appl
45	307	4.3	726	4	US-10-282-122A-72033	Sequence 72033, A

ALIGNMENTS

RESULT 1
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 100.0%; Score 7143; DB 3; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MEKKVRFKLKRVKRWTVTSIASAVVTLTSLGSLVKADSTDDRQAVTESQASLVTTSE	60
Db	1	MEKKVRFKLKRVKRWTVTSIASAVVTLTSLGSLVKADSTDDRQAVTESQASLVTTSE	60
QY	61	AAKETLTATDTSATSQPTATVTDNVSTNTNTANTANFVVKPTTSEQAKTDN	120
Db	61	AAKETLTATDTSATSQPTATVTDNVSTNTNTANTANFVVKPTTSEQAKTDN	120
QY	121	SDKIITTSKANRLTATCKFVPANNNTAHPKVTVDKIPIKPKICKLPSSLSODDIAA	180
Db	121	SDKIITTSKANRLTATCKFVPANNNTAHPKVTVDKIPIKPKICKLPSSLSODDIAA	180
QY	181	LGNVKNTRKVGKYYKEDGTLOQNVALNINGKTFFFDEYTGALSNNTLPSKKGNITNND	240
Db	181	LGNVKNTRKVGKYYKEDGTLOQNVALNINGKTFFFDEYTGALSNNTLPSKKGNITNND	240

Db 181 LGNVNIRKVGKYYKEDGTLQKNYALNINGKTFPFDGTGALSNNLTLPSSKGNITNND 240
Qy 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEISWYRPKYILKDGKTWTQSTEKDPRPLMTW 300
Db 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEISWYRPKYILKDGKTWTQSTEKDPRPLMTW 300
Qy 301 PQDETQRYVVMNAQLGIHQTYNTATSPQLNLAAQTIOTKIEEKI TAEKNTNMLRQTI 360
Db 301 PQDETQRYVVMNAQLGIHQTYNTATSPQLNLAAQTIOTKIEEKI TAEKNTNMLRQTI 360
Qy 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQSANSYRILNRTPTNQTGKQDPR 420
Db 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQSANSYRILNRTPTNQTGKQDPR 420
Qy 421 YTADRTIGGYEFLANDVNSNPVQAEQLNLHFLMNFNGNTYANDPDANFDSIRVDAVD 480
Db 421 YTADRTIGGYEFLANDVNSNPVQAEQLNLHFLMNFNGNTYANDPDANFDSIRVDAVD 480
Qy 481 NVADLLQIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDTPYLHDDGDNNMNDNRLR 540
Db 481 NVADLLQIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDTPYLHDDGDNNMNDNRLR 540
Qy 541 LSLLYSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVSYSFIRAHDSVODLIRNIIR 600
Db 541 LSLLYSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVSYSFIRAHDSVODLIRNIIR 600
Qy 601 TEINPNVGYSTFTEIKAFEIYNKDLLATEKKYTHYNTALSYALLTNKSSVPRVYVG 660
Db 601 TEINPNVGYSTFTEIKAFEIYNKDLLATEKKYTHYNTALSYALLTNKSSVPRVYVG 660
Qy 661 DMFTDDGQYMAKTIINYEAIETLLKARIKYVSGQAMRNQOVGNSEIITSVRYGKALKA 720
Db 661 DMFTDDGQYMAKTIINYEAIETLLKARIKYVSGQAMRNQOVGNSEIITSVRYGKALKA 720
Qy 721 TDTGDRTRTSGVAVIEGNPNLSRLKASDRVVVNMGAHKNOAYRPLLLTTDNGIKAYHS 780
Db 721 TDTGDRTRTSGVAVIEGNPNLSRLKASDRVVVNMGAHKNOAYRPLLLTTDNGIKAYHS 780
Qy 781 DQEAAGLVRYTNDRGELIPTAADIKGYANPQVSGYLGWVPVGAADQDVRVAASTAPST 840
Db 781 DQEAAGLVRYTNDRGELIPTAADIKGYANPQVSGYLGWVPVGAADQDVRVAASTAPST 840
Qy 841 DKSQVHQNAAALSRVWFEGFSNFQAFATKKEEYTNVVIKAKNDVKAEMGVTDPEWAPQV 900
Db 841 DKSQVHQNAAALSRVWFEGFSNFQAFATKKEEYTNVVIKAKNDVKAEMGVTDPEWAPQV 900
Qy 901 SSTDGSFLDSVIQNGYAFTRDYDLGISKENKYGTDADLVKAIKALHSGIKYMWADVPDQ 960
Db 901 SSTDGSFLDSVIQNGYAFTRDYDLGISKENKYGTDADLVKAIKALHSGIKYMWADVPDQ 960
Qy 961 MYALPKEVVTATRVDKYGTVPAGSQIKNTLVVDGKSSGKQQAQYGGAFLEELQAKYP 1020
Db 961 MYALPKEVVTATRVDKYGTVPAGSQIKNTLVVDGKSSGKQQAQYGGAFLEELQAKYP 1020
Qy 1021 ELFPARKQISTGVPMDPSVKIKOWSAKYFNGTNILGRGAGYVLKDAQNTYFSLVSDNTP 1080
Db 1021 ELFPARKQISTGVPMDPSVKIKOWSAKYFNGTNILGRGAGYVLKDAQNTYFSLVSDNTP 1080
Qy 1081 PKSLVNPNHGTSSTVGLVFDGKGYVYYSYTSNGQAKNAFISLGNWNYFDNNGYVMTGAQ 1140
Db 1081 PKSLVNPNHGTSSTVGLVFDGKGYVYYSYTSNGQAKNAFISLGNWNYFDNNGYVMTGAQ 1140
Qy 1141 SINGANYTPLSNGIQLRNAIYONGNKVLSYNGDGRYENGYYLFGQQWRYFQNGIMAVG 1200
Db 1141 SINGANYTPLSNGIQLRNAIYONGNKVLSYNGDGRYENGYYLFGQQWRYFQNGIMAVG 1200
Qy 1201 LTRVHGAVQYFDASGFQAKGFITTTADGKLYFDRDSGNQISNRRFVRNSKGWELFDHNG 1260
Db 1201 LTRVHGAVQYFDASGFQAKGFITTTADGKLYFDRDSGNQISNRRFVRNSKGWELFDHNG 1260
Qy 1261 VAVTGTVTENGQRLYFKPKNQVQAKGEFIRDANGYLRYDPPNSGNEVRNRFVRNSKGWEL 1320
Db 1261 VAVTGTVTENGQRLYFKPKNQVQAKGEFIRDANGYLRYDPPNSGNEVRNRFVRNSKGWEL 1320

RESULT 2

US-10-383-930-35
; Sequence 35, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-35

Query Match 100.0%; Score 7143; DB 4; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEKKVRFKLRKVKRWVTVSIAASAVVTLTSLSGSLVKADSTDDRQAAVTESQASLVTTSE 60
Db 1 MEKKVRFKLRKVKRWVTVSIAASAVVTLTSLSGSLVKADSTDDRQAAVTESQASLVTTSE 60
Qy 61 AAKETLTATDSTATSATSQPTATVTDNVSTNQSTNTTANTANFVVKPTTTTSEQAKTDN 120
Db 61 AAKETLTATDSTATSATSQPTATVTDNVSTNQSTNTTANTANFVVKPTTTTSEQAKTDN 120
Qy 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKI VIPKPKIGKLQPSLSQDDIAA 180
Db 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKI VIPKPKIGKLQPSLSQDDIAA 180
Qy 181 LGNVNIRKVGKYYKEDGTLQKNYALNINGKTFPFDGTGALSNNLTLPSSKGNITNND 240
Db 181 LGNVNIRKVGKYYKEDGTLQKNYALNINGKTFPFDGTGALSNNLTLPSSKGNITNND 240
Qy 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEISWYRPKYILKDGKTWTQSTEKDPRPLMTW 300
Db 241 NTNSFAQYNQVSTVDVANFEHVDHYLTAEISWYRPKYILKDGKTWTQSTEKDPRPLMTW 300
Qy 301 PQDETQRYVVMNAQLGIHQTYNTATSPQLNLAAQTIOTKIEEKI TAEKNTNMLRQTI 360
Db 301 PQDETQRYVVMNAQLGIHQTYNTATSPQLNLAAQTIOTKIEEKI TAEKNTNMLRQTI 360
Qy 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQSANSYRILNRTPTNQTGKQDPR 420
Db 361 SAFVKTQSAWNSDSEKPPDDHLQKALLYSNNSKLTQSANSYRILNRTPTNQTGKQDPR 420
Qy 421 YTADRTIGGYEFLANDVNSNPVQAEQLNLHFLMNFNGNTYANDPDANFDSIRVDAVD 480
Db 421 YTADRTIGGYEFLANDVNSNPVQAEQLNLHFLMNFNGNTYANDPDANFDSIRVDAVD 480
Qy 481 NVADLLQIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDTPYLHDDGDNNMNDNRLR 540
Db 481 NVADLLQIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDTPYLHDDGDNNMNDNRLR 540
Qy 541 LSLLYSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVSYSFIRAHDSVODLIRNIIR 600
Db 541 LSLLYSLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVSYSFIRAHDSVODLIRNIIR 600

QY 601 TEINPNVVGYSFTTBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKNSVPRVYIG 660
Db 601 TEINPNVVGYSFTTBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKNSVPRVYIG 660
QY 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALK 720
Db 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALK 720
QY 721 TDGDRTRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKNAQYRPLLLTTDNGIKAYHS 780
Db 721 TDGDRTRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKNAQYRPLLLTTDNGIKAYHS 780
QY 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASAPST 840
Db 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASAPST 840
QY 841 DGKSVHQNAAALDSRVNMFEGFSNFOAFATKKEEYTNVVIKNVDKPAEWGVTDFEMAPQYV 900
Db 841 DGKSVHQNAAALDSRVNMFEGFSNFOAFATKKEEYTNVVIKNVDKPAEWGVTDFEMAPQYV 900
QY 901 SSTDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
Db 901 SSTDGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQ 960
QY 961 MYALPEKEVVTATRVVDKYGTPVAGSQIKNTLYYVDGKSSGKQQAQYGGAFLEELQAKYP 1020
Db 961 MYALPEKEVVTATRVVDKYGTPVAGSQIKNTLYYVDGKSSGKQQAQYGGAFLEELQAKYP 1020
QY 1021 ELFPARKQISTGVPMDBSVKIKOWSAKYFNGTNILGRGAGVYLKQDQNTNYFSLVSDNTFL 1080
Db 1021 ELFPARKQISTGVPMDBSVKIKOWSAKYFNGTNILGRGAGVYLKQDQNTNYFSLVSDNTFL 1080
QY 1081 PKSLVNPNHGTSSTGLVFDGKGYVYYSYSGNQAKNAFISLGNWYDFDNGYMTGAQ 1140
Db 1081 PKSLVNPNHGTSSTGLVFDGKGYVYYSYSGNQAKNAFISLGNWYDFDNGYMTGAQ 1140
QY 1141 SINGANYPLSNGIQLRNAIYDNGKNVLSYNGDGRYENGYYLFCQWRYFQNGIMAVG 1200
Db 1141 SINGANYPLSNGIQLRNAIYDNGKNVLSYNGDGRYENGYYLFCQWRYFQNGIMAVG 1200
QY 1201 LTRVHGAVOYFASGFOAQGFITTDAGKLRVFDKSGNOISNRFRVNSKGWFLPDHNG 1260
Db 1201 LTRVHGAVOYFASGFOAQGFITTDAGKLRVFDKSGNOISNRFRVNSKGWFLPDHNG 1260
QY 1261 VAVTGTVTFNGORLPFKPQNGVQAKGEFIRDANGYLRYPDPNSGNEVRNRFVNSKGWFL 1320
Db 1261 VAVTGTVTFNGORLPFKPQNGVQAKGEFIRDANGYLRYPDPNSGNEVRNRFVNSKGWFL 1320
QY 1321 FDHNGIATVGARVNGHASILSLMVPRLESSLQSVKVVSNMTMILIPMKFVIVM 1375
Db 1321 FDHNGIATVGARVNGHASILSLMVPRLESSLQSVKVVSNMTMILIPMKFVIVM 1375

RESULT 3
US-10-797-821-35

; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550

; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-35

Query Match 100.0%; Score 7143; DB 5; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKKVRPKLRKVKRWVTVSIAASVVTLSLSGLVKADSTDRQQAQVTSQASLVTTSE 60
Db 1 MEKKVRPKLRKVKRWVTVSIAASVVTLSLSGLVKADSTDRQQAQVTSQASLVTTSE 60
QY 61 AAKETLTATDTSTATSQPTATVTDVNSTNQSTNTTANTANFVVKPTTSEQAKTDN 120
Db 61 AAKETLTATDTSTATSQPTATVTDVNSTNQSTNTTANTANFVVKPTTSEQAKTDN 120
QY 121 SKLIITTSKAVNRLTATGKFPVANNNTAHPKVTVDKI VPIKPIKIGLKQPSSLSQDDIAA 180
Db 121 SKLIITTSKAVNRLTATGKFPVANNNTAHPKVTVDKI VPIKPIKIGLKQPSSLSQDDIAA 180
QY 181 LGNVNIRKVGKYYKYKEDGTLQKNYALNINIKTEFFDETGALSNTLPSKKGNITNND 240
Db 181 LGNVNIRKVGKYYKYKEDGTLQKNYALNINIKTEFFDETGALSNTLPSKKGNITNND 240
QY 241 NTNSFAQYNQVSTVDVANFEVDHVLTAESWYRPKYILKDGKTWTQSTEKDFPLMTW 300
Db 241 NTNSFAQYNQVSTVDVANFEVDHVLTAESWYRPKYILKDGKTWTQSTEKDFPLMTW 300
QY 301 PQDETQRQVYVNMNAQLGIHQTYNTATSPLOLNLAQTTIQTKEEKITAEKNTNWLRTI 360
Db 301 PQDETQRQVYVNMNAQLGIHQTYNTATSPLOLNLAQTTIQTKEEKITAEKNTNWLRTI 360
QY 361 SAFVKTOSAWNSDSKPPDDHLQKALLYSNNSKLTQSOANSYRILNRPPTNQTGKKDPR 420
Db 361 SAFVKTOSAWNSDSKPPDDHLQKALLYSNNSKLTQSOANSYRILNRPPTNQTGKKDPR 420
QY 421 YTADRTIGYEYFLLANDVNSNPVQAEQLNHLFLMNFNGNIYANDPDANFDSIRVDADV 480
Db 421 YTADRTIGYEYFLLANDVNSNPVQAEQLNHLFLMNFNGNIYANDPDANFDSIRVDADV 480
QY 481 NYDADLLQIAGDYLLKAAKGIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNMINMDNRLR 540
Db 481 NYDADLLQIAGDYLLKAAKGIHKNDKAAANDHLSILEAWSYNDTPYLHDDGDNMINMDNRLR 540
QY 541 LSLLYSLAKPLNORSQGNPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVODLIRNIIR 600
Db 541 LSLLYSLAKPLNORSQGNPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVODLIRNIIR 600
QY 601 TEINPNVVGYSFTTBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKNSVPRVYIG 660
Db 601 TEINPNVVGYSFTTBEIKKAFIYNKOLLATEKKYTHYNTALSYALLLTKNSVPRVYIG 660
QY 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALK 720
Db 661 DMFTDDGQYMAHKTINYEAIETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALK 720
QY 721 TDGDRTRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKNAQYRPLLLTTDNGIKAYHS 780
Db 721 TDGDRTRTSGVAVIEGNNPSLRLKASDRVVVNMGAHKNAQYRPLLLTTDNGIKAYHS 780
QY 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASAPST 840
Db 781 DQEAAGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVPVGAADQDVRVAASAPST 840
QY 841 DGKSVHQNAAALDSRVNMFEGFSNFOAFATKKEEYTNVVIKNVDKPAEWGVTDFEMAPQYV 900


```
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRP
; ORGANISM: Streptococcus mutans
; US-10-797-821-34

Query Match          74.0%; Score 5285; DB 5; Length 1475;
Best Local Similarity 75.9%; Pred. No. 1.1e-296;
Matches 1031; Conservative 101; Mismatches 169; Indels 58; Gaps 14;

QY 1 MEKKVRFKLVKRWVTSIASAVVLTLSGSLVKADSTDRQQAUTESQASLVTTSE 60
DB 1 MDKVRVYKLVKRWVTSVASAVMTLTLSGGLVKADSNESKSQISNDSNTSVVTANE 60
QY 61 AKKTLTATDSTATSQPTA--TVTDNSTTQSTNTTANTANFVVKPTTSEQAKT 118
DB 61 -----ESNVIETATSKQEAASQTNHTVTTSSSTSVVNPKVSVNPFYTVGETA-- 109
QY 119 DNSDKIITTSKAVNLTATGKVPANNNTAHPKTVTDKIVPIKPKIGLKQPSLSQDDI 178
DB 110 SNGEKLQNTTVDKTSQA-----AANNISKQTTED-----TDVIDSN 149
QY 179 AA----LGNVKNIRKNGKYYYKEDGTLQKNYALNINNGKTFPDETGALSNNTLPKKG 234
DB 150 AANLQILEKLPNVKIDGKYVYDNGKVRNFTLIADGKILHFDGTGAYTDSIDTVNK 209
QY 235 NITNDNTNSPAQYNOVSTVDVANEHVDHYLTAEVWRPKYILKDGKTWTQSTKDRP 294
DB 210 DIVTT-RSNLYKYNQVDRSAQSFEHVDHYLTAEVWRPKYILKDGKTWTQSTKDRP 268
QY 295 LLMTWPPQETQOVVYVYNAQLGHTQVNTATSPQLNLAQTITQIEEKITAEKNTN 354
DB 269 LLMTWPPQETQOVVYVYNAQLGINKTYDDTSNQLNLNAAATTOAKIEAKITLKNTD 328
QY 355 WLQRTISAFVKTQSAWNSDSEKPFDDHLQKGLLYSNNSKLTQSAANSYRILNRTPTNQT 414
DB 329 WLQRTISAFVKTQSAWNSDSEKPFDDHLQNGAVLYDNEGKLTPTYANSYRILNRTPTNQT 388
QY 415 GKQDPRTYADRTIGGYEFLLANDVNSNPVQAEQNLWLHFLMNFNGNIYANDPDANFDSI 474
DB 389 GKQDPRTYADNTIGGYEFLLANDVNSNPVQAEQNLWLHFLMNFNGNIYANDPDANFDSI 448
QY 475 RVDADVNDVADLLQIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDTPYLLHDDGNNIN 534
DB 449 RVDADVNDVADLLQIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDTPYLLHDDGNNIN 508
QY 535 MDNRLRLSLLSLAKPLNQRSGMNPPLITNSLVNRTDDNAETAAPVPSYGFIRAHDSVQDL 594
DB 509 MDNKLRLSLLSLAKPLNQRSGMNPPLITNSLVNRTDDNAETAAPVPSYGFIRAHDSVQDL 568
QY 595 IRNIIRTEINPNVGYSTTEIEKAPFIYNKDLATEKKYTHYNTALSYALLLNKSSV 654
DB 569 TADIIKAEINPNVGYSTTEIEKAPFIYNKDLATEKKYTHYNTALSYALLLNKSSV 628
QY 655 PRVYVYDGMFTDDQYMAHKTINYEAEITLLKARIKYVSGQAMRNQOVGNSBIITSVRYG 714
DB 628 PRVYVYDGMFTDDQYMAHKTINYEAEITLLKARIKYVSGQAMRNQOVGNSBIITSVRYG 688
QY 715 KGALKATDTGDRTRTTSVAVIEGNNPSLRLLKASDRVVVNMCAAHKNOAYRPLLLTTDNG 774
DB 689 KGALKATDTGDRTRTTSVAVIEGNNPSLRLLKASDRVVVNMCAAHKNOAYRPLLLTTDNG 748
QY 775 IKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYNANPOVSGYLVGVVVPVGAALIKWFALE 808
DB 749 IKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYNANPOVSGYLVGVVVPVGAALIKWFALE 808
QY 833 AASTAPSTDGKSVHQNAAALDSRVMPFESNFQAFATKKEEYTNVVIKNDVFAEWGVTD 892
DB 809 A---RPHQOMASVHQNAAALDSRVMPFESNFQAFATKKEEYTNVVIKNDVFAEWGVTD 865
QY 893 FEMAPQYVSSSTDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIGV 952
DB 866 FEMAPQYVSSSTDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIGV 925
QY 953 MADVVPDQMYALPEKEVVVATRVKYGTPVAGSQIKNTLYVVDGKSSGKDDQAKYGGAPL 1012
DB 926 MADVVPDQMYALPEKEVVVATRVKYGTPVAGSQIKNTLYVVDGKSSGKDDQAKYGGAPL 985
QY 1013 EELQAKYPELFARKQISTGVPMDFSVKIKQWSAKYFNGTNIILGRGAGYVVKDQATNTYPS 1072
DB 986 EELQAKYPELFARKQISTGVPMDFSVKIKQWSAKYFNGTNIILGRGAGYVVKDQATNTYPS 1045
QY 1073 LVSDN---TFPLPKSLVNPNGHTSSVTGLVFDGKGYVYVYSTSGNOAKNAFISLGNWYIF 1129
DB 1046 -ISDNKEINFLPKTLN-----QDSQVGFSDYDGKGYVYVYSTSGYQAKNTFISEGDKWYIF 1099
QY 1130 DNGVMVTGAOSINGANYFELSNGLQRLNAYLDNCKNLYVYVYDNGRRYENGYIIF-GQQ 1188
DB 1100 DNGVMVTGAOSINGANYFELSNGLQRLNAYLDNCKNLYVYVYDNGRRYENGYIIF-GQQ 1159
QY 1189 WRYFQNGIMAYGLTRVHGAVQYFDASGFAQKQFITTADGKLYRFDSDSGNOISNRFVN 1248
DB 1160 WRHFNNGEMSVGLTVIDGQVQYFDEMVGQAKGKFTVTADGKIRYFDKQSGNMYRNRFIN 1219
QY 1249 SKGEWFLFDHNGVAVTGTFTNGQRLYFKPNQVQAKGEFIRDANGYLYRYDPSNGNEVN 1308
DB 1220 BEGKWLVLGEDGAAVGTSGTNGHLYFRANGVQVKGFEVTDHGHGRISYDNGSGDQIRN 1279
QY 1309 RPYRNSKGEFLFDHNGIAVTGARVNVGHASITLSLMVFR 1347
DB 1280 RFVRNAQQWFFYFDNNGYAVTGAARTING-----QLLYFR 1313

RESULT 7
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRP
; ORGANISM: Streptococcus sobrinus
; US-10-383-930-37

Query Match          59.0%; Score 4214.5; DB 4; Length 1590;
Best Local Similarity 60.3%; Pred. No. 1e-234;
Matches 811; Conservative 186; Mismatches 296; Indels 53; Gaps 14;
```


Db	437	QLNWHLNYGSI	VANDPEANFDGVR	VDVNDVNADLLQ	IASDYLKAHYGVDS	SEKNAI	496
Qy	509	DHLSILEAWSN	DTPYLHDDGD	GNMINMDNR	LRLSLLYSLAKP	LNQ-----RSGM	NPL 560
Db	497	NHLSILEAWS	DNDPOYNKDT	KGAQLPIDN	KLRLSILYALT	TRPLEKASNKNEI	RSGLPEV 556
Qy	561	ITNSLWRTD	DDNAETAAP	VSFSFTRAHD	SEVQDILLIRNI	IRTEINPNVV	GYSFTEETKKA 620
Db	557	ITNSLNNRS	ABGKNSERM	ANITFIRAHD	SEVQTVIAKI	IKAQINPKTD	LGLTFTLDELKQA 616
Qy	621	FEIYNKOLL	ATTEKKYTH	NTALSTALS	VALLATNKSS	VPVYGVDMPT	DDGOYMAHKNTINEAI 680
Db	617	FKIYNED	MURQAKKT	QTSNITPA	TALMUSNKDS	ITRLYIGDMS	DDGOYMATKSPYDAI 676
Qy	681	ETLLKARI	KIYVSGGOA	MNQGVNSE-----	IITSVRYG	KGALKATD	TGDRTRTS 731
Db	677	DTLLKARI	KIYAAGQ	DMKITTYEGD	KSHMDWDY	TGVLTSVRYG	TGANEATDQSEAKTKQ 736
Qy	732	GVAITEG	NPSLRILK	ASDRVVVN	MGAAHKNQ	AYRPLLT	TTDONGIKAYHSDQEAAGLVRYT 791
Db	737	GMAVITSN	NPISKL	QNQDKVT	VNMGAAHKNQ	EYRPLLT	TTKDGULTSVTSDAAAKSLVRKT 796
Qy	792	NDRGELI	FTADI	IKGYAN	POVSYGLV	WVPVGAAD	QDVRVAASTASTDCKSVHQAAL 851
Db	797	NDKGEL	VEDASDI	OGYLP	POVSGYLA	VWVPVGS	DNQDQDVRVAASNKANATQVYESSAL 856
Qy	852	DSRVW	FEGFSNFQ	AFATKKE	BYTNVIAKN	VDKFAEWG	VTDFEMAPOVVSSTDCGSFLDQSV 911
Db	857	DSQLI	YEGFSNFQ	DFVTKDS	YTNKKIAQ	NVOLFKS	WGVTFSFEMAPQVVSSEDCGSFLDQSI 916
Qy	912	IQNGYA	FTRDYDL	GIS	PNKTYGTADD	LVKAIKAL	HSKGIKVMADWPQDMYALPEKEVVT 971
Db	917	IQNGYA	FEDRYDL	AMS	KNKYGSQ	ODMINAV	KALHKSIGIOIADWVPDQIYNLPGKEVVT 976
Qy	972	ATTRVD	KYTPVAG	SQIKNTL	VYVDGK	SGKDOQA	KYGAFLELQAKYPFLPARKQISTG 1031
Db	977	ATTRVND	GYSTRK	DBSEIKNTL	YAANTK	NGKDDYQ	AKYGAFLELAAKYPISFNRTQISNG 1036
Qy	1032	VPMDPS	VKIKQWS	AKYFNG	TNILGR	GAGYVLK	QOATNTYFSLVSDNTFLPKSLVNPNGHT 1091
Db	1037	KKIDP	SEKITA	WKA	KYFNG	TNILGR	GVGYVLKONASKEYFLKGNQYILPKQMTN----- 1091
Qy	1092	SSSVT	GLVFDG	KGYVY	YSTG	NOAKNAFI	-SLGNWNYFDNNGYMTGASINGANYFFL 1150
Db	1092	KEASTG	FVNDG	MGTFTYST	GYQAKNS	FVQDAG	KNWYIFDNNGHMVYGLQLNGEVQYFL 1151
Qy	1151	SNGLQ	LRNAYLD	NGNKVLS	YVNDGR	YENGYYLF--	GOQWRYPQ-NGIMAVGLTRVHGA 1207
Db	1152	SNGVQL	RESFLEN	ADGSKYF	GHILGN	YSNGYYS	FSDNDSKWRYPDASGVNMAVGLKTINGN 1211
Qy	1208	VOYPDAS	GFOAKG	OFITAD	GKLRYFP	DRDSNQI	SNRFVNSKGEWFLFDHNGVAVTGTV 1267
Db	1212	TQYFDQ	DGYQV	KGAWIT	YTGSDG	KGYFDG	SGNMAVNRFPANDKNGDWYLLNSDGLALVGVO 1271
Qy	1268	TFNGQ	RLYFKP	NGVQAK	GEFTR	DANGY	LYRYDDPNSNGEVNRNRFVNSKGSWFLFDHNGIA 1327
Db	1272	TINGK	TYFQD	GGQIK	GKGIITD	-NGKLKY	FLANSGLARNIATFDSQNNWYFYGSDGVA 1330
Qy	1328	VTGAR	VVNG	1336			
Db	1331	VTGS	OTIAG	1339			

RESULT 10

RESULT 10
US-10-383-930-36

US-10-363-930-38
; Sequence 36, Application US/10383930

; sequence 36, APPLICATION US/103
 ; Publication No. US20040127400A1

; PUBLICATION NO: US20
; GENERAL INFORMATION:

APPLICANT: Smith, Daniel J

APPLICANT: SMITH, DAVID J
APPLICANT: Taubman, Martin A

APPLICANT: IAUHMAN, MARTIN A
TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

; TITLE OF INVENTION: IMMUNE
; FILE REFERENCE: 25669-018

FILE REFERENCE: 25669-018
CURRENT APPLICATION NUMBER: US/10/383,930

RESULT 13
US-10-797-821-38
; Sequence 38, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Tauhman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-38

Query Match 44.3%; Score 3161.5; DB 5; Length 1554;
Best Local Similarity 48.5%; Pred. No. 7.8e-174;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;

QY 1 MEKKVREKLKVKRWVTVSIAAVVTLTSLGSLVKADSTDRQQAATESQASL----- 55
DB 1 MEKKVREKLKVKRWVTVSIAAVVTLTSLGSLVKADSTDRQQAATESQASL----- 55
QY 56 ---VTTSEAAKTLTATDTSTATSQTATVTDNVSTNTQSTNTTANTANFVVKPTTT 112
DB 57 QDQATTNDANTWT-DTDTADQSANINQDQAG-SQSNQDQAKQDQNT-----DRNQ 108
QY 113 SEQAKTDSNDKIITTSKAVNRLTATGKFPVANNNTAHPKTVDKVPKPKGKUKQPS 172
DB 109 ADNSQTDN-----NQATDQATS-----PATDGTSVQRDAANVATAADQEG--QTAP 153
QY 173 LSQDDIAALGNVKNIRKYNKYKYYKEDGTLQKNVALNINGKTFPPD-ETGALSNTLPS 231
DB 154 SEQEKSAAL-SLDNVKLDGKYIYVQADGSKYKKNFAITVNGQMLYFSDTGALSTSTYS 212
QY 232 KKGNTTNDNTNSFAQYNQVSTVDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTED 291
DB 213 FSQGTN--LVDDFSSHKNAYDSTAKSFELVNGYLTANSWYEPAGILRNGQTWEASNEND 270
QY 292 FRPLMTWMPDQETQRYVYNNALQIHTYNT-ATSPQLNLAAQTQIKIEKITAE 350
DB 271 LRPVLMSWMPDKDTQVAYVYNNKYLANSATEVTNETSQVDLKEAQSIQTQIEKITSD 330
QY 351 KNTNWLRTQISAFVKTQSAWNSDSEKPPD--DHLQKGLLYSNNSKLTQSAWNSYRLNR 408
DB 331 NSTQWLRTAMEAFVAAQPKWNMSTEN-FNKDHLQGLGALLYT-NSDLTPWANSYRLNR 388
QY 409 TPTNQTGKDPRTYADRTGGYEFLLANDVNSPNVQAEQLNLHLFLNFGNIYANDPD 468
DB 389 TPTQDQGT--KYFTEGEGGEGYFLLSNDVNSPNVQAEQLNLHLNMGDIYWGDKD 446
QY 469 ANFDSIRVDVNDVADLLQIAGDYLKAAKGHKNDKANDHLSILEAWSYNDTPLYHDD 528
DB 447 ANFDGVRVDVNDVNDVADLLQVYSNYFKONYKYVTDSANALAHISILEAWSLNDQYNET 506
QY 529 GDNMNMNRLRLSLLSLAKPLNQRSGMNPILITNSL-VNRDNDNAETAAPVPSYFIRAH 587

Db 507 NGTALSIDNSSRLTSLAVLTQPGQRIDLSNLSVSKERANDTAYGDTITFYFVRAH 566
QY 588 DSEVODLIRNIIRTEINPNVGVSTPTEIIEKAPFIYNKDLATEKKYTHYNTALSAYLL 647
Db 567 DSEVQTVIAKIVKEKIDTNSDGYTFLDQLKDAFKIYNEDMAKVNKTTHYNIAPAYLL 626
QY 648 LTNKSSPRVYVYVGMFTDDGQYMAHKTINYEAIEFLKARIKYVSGGQAMRNQOV-GNSE 706
Db 637 LSNMESPRVYVYVGLYTDGQYMAKSPYDAIATMLQGRIAIYVSGGSEEVHKVGNNGNQ 686
QY 707 IITSVRYGKALKATDT-GDRTTRTSGVAVTEGNNPSLRKASDRVVMGAHKNQAYR 765
Db 687 ILSSVRYGQDLMSADDTQGTDLSTRTSLVTLVNSDPNLDL-GGDSLTVMGHAHQAAYR 745
QY 766 PLLLTDDNGIKAYHSDQBAAGLVRYTNDRGELIFTAADIKGYANPQVGYLGVVPGVAA 825
Db 746 PLLILGTDGVQSYLKDS-D-TNIVKYTDANGNLTFADDIKGYSTVDMGYLAVVYPVGA 804
QY 826 ADQVRAASTAPSTDGKSVHQNALDSRVNPEGESNFQAFATKKEEYTNVVIKNDKF 885
Db 805 DQDQVRAADTNQKADGSKSLKTSAAALDSQVIEGFSNFQDFANNDADYTNKKIADNADFF 864
QY 886 AEWGVTDPFEMAPQVYSSDGSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKAL 945
Db 865 KKLGITSEMAPQVYSATDGSFLDSIIQNGYAFSDRYDLAMSKNKYGSKDDLAKALKAL 924
QY 946 HSKGIKVMADWVPDQMYALPEKEVVVATRVKYGTPPVAGSQIKNTLYVVDGKSSGKQQA 1005
Db 925 HANGIQAIDWVPDQIYQLPGEEVVVATKRTNSYGNPTPDAYINNALLYATNTKSSGSDYQA 984
QY 1006 KYGGAFLLELOAKYPELPFARKQISTGVPMDSVKIKOWSAKYFNGTNTILGRGAGVYLKQ 1065
Db 985 QYGGAFDELKAKYPMDFVNMISTGKPIDPSTKIKOWEAKYFNGTNTVLGKAGVYLSDD 1044
QY 1066 ATNTYFSLVSDNTLPLKVLNPNHGTSSVGLVDFGKGVYVYVYSGNOAKNAFISLGN 1125
Db 1045 ATGKYFTVNGDFLPSFT----GDQNAKTGYFDGTGMAYYSTSGKAVNSFYEGH 1100
QY 1126 WYFDNNGYMTGA-QSINGANYPLSNGIQLRNALYDNGNKVLSYNGDGRYRNGYVL 1184
Db 1101 YYYFDKGMVTSYKABDGNDDYFLPNGIQMRDAIYQDAQNSYYYGRTGLYK----- 1155
QY 1185 FGQW-----RYFO-NGINAVGLTRVHGAVOVYFDSGFOAKQOFITADGK 1229
Db 1156 -GDMNYPFVDPNNANKTVFRYFDANNVMAIGYRNMYGQTYYPDENGFOAKGOLLTDDKG- 1213
QY 1230 LRYFDRDSGNOISNRFRVNSKGEWFLFDHNGVAVTGTVTTFNGQRLYFKP-NGVQAKGEFI 1288
Db 1214 THYFEDNGAMAKNFV-NVGDDWYMDGNGNAVKGQYVNNQILLYFNPETGVQVKGQFI 1272
QY 1289 RDANGLYRYDPNCGNEVRNFRVNSKGEWFLFDHNGIAVTGARVV 1334
Db 1273 TDAQRTSYDANSALKSSGFFTPNGSDWY-YAENGYYVYKGFQV 1317

RESULT 14
US-10-484-218-18
; Sequence 18, Application US/10484218
; Publication No. US20050059633A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; TITLE OF INVENTION: LACTIC ACID BACTERIA
; FILE REFERENCE: 2001-1316
; CURRENT APPLICATION NUMBER: US/10/484,218
; CURRENT FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: PCT/NL02/00495
; PRIOR FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: EP 01202752.0
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: EP 01202841.1
; PRIOR FILING DATE: 2001-07-25

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 7.82456 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-35
Perfect score: 7143
Sequence: 1 MEKKVRFKLRVKVGRVTVS.....VKVNSVNTMLIPMKFVIVM 1375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	308.5	4.3	2710	US-11-051-453-41	Sequence 41, Appl
2	307	4.3	726	US-11-052-554A-217	Sequence 217, App
3	266	3.7	1155	US-10-793-626-1780	Sequence 1780, Ap
4	255.5	3.6	2367	US-11-051-453-42	Sequence 42, Appl
5	251.5	3.5	2314	US-11-013-759-11	Sequence 11, Appl
6	245.5	3.4	1992	US-11-013-759-3	Sequence 3, Appl
7	245.5	3.4	1992	US-11-013-759-13	Sequence 13, Appl
8	245.5	3.4	2047	US-11-013-759-4	Sequence 4, Appl
9	245.5	3.4	2047	US-11-013-759-7	Sequence 7, Appl
10	235	3.3	701	US-11-052-554A-231	Sequence 231, App
11	232.5	3.3	619	US-11-052-554A-229	Sequence 229, App
12	227	3.2	690	US-11-052-554A-232	Sequence 232, App
13	227	3.2	2902	US-11-052-554A-91	Sequence 91, Appl
14	223.5	3.1	1647	US-11-052-554A-260	Sequence 260, App
15	223	3.1	5291	US-11-052-554A-281	Sequence 281, App
16	222	3.1	2399	US-11-052-554A-92	Sequence 92, Appl
17	220	3.1	1588	US-11-052-554A-280	Sequence 280, App
18	217	3.0	396	US-11-022-562-228	Sequence 228, App
19	217	3.0	2053	US-11-013-759-9	Sequence 9, Appl
20	214	3.0	1237	US-11-052-554A-95	Sequence 95, Appl
21	213	3.0	5024	US-10-793-626-2964	Sequence 2964, Ap
22	212.5	3.0	483	US-10-630-203-8	Sequence 8, Appl
23	212.5	3.0	483	US-10-630-203-30	Sequence 30, Appl
24	212.5	3.0	512	US-11-102-188-5	Sequence 5, Appl
25	212	3.0	483	US-10-630-203-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-FEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 4.3%; Score 308.5; DB 7; Length 2710;
Best Local Similarity 20.4%; Fred. No. 2.5e-10;
Matches 321; Conservative 194; Mismatches 542; Indels 517; Gaps 82;
QY 92 TNSQNTTANTANFVVKPTTSE-----QAQTD-- 119
DB 1286 TNIKLDKDTNFIMPITITNEIRNKLSYSGDGGTYSLLSSYPITNINLSKDDLM 1345
QY 120 --NSDKITTSKAVNRLPATGKFPANNNAHPKTVTDKIVPIKPKIGLKQPSSLSQDD 177
DB 1346 IFNIDNEVREISIEGTTIKGKLI-----KDVLSKIDINKNL-----IIGNOT 1389
QY 178 IAAIGNVQN-----IRKVGKYYYKEDGTLOKNYALNINGKTFPFDETCALSNNTLP 230
DB 1390 IDFGSDINDKORYIFLTCELDKISLIITEINLVAKSYSLLLSG-----DKNYLISNLSNT 1444

```
QY 231 SKKGNITNDNTNSPAQVNVYSTVDANFEHVDHYLTAEWSYRPKYIL---KDGKT--- 283
Db 1445 IEKINTLGLDSKNIAYV---TBSN---NKTFGAISKTSQSIHYHKDQSKNILEF 1495
QY 284 WTQST---EKDF---RPLMTWPDQETQ--- 307
Db 1496 YNDSTLEFNSKDFIAEDINVMKDDINTITGKYVYVDDNNTDKSIDFSISLVSKNQVKVNGL 1555
QY 308 ---QYVYMAQGIHQTYNTATSPLOLNAAQTIQIKBEKITAENKTNWLRQT 359
Db 1556 YLVNESYSSYLDFFVKNSDGHNTSF---MNLFLONISF---WKLFGPENINFIIDK 1606
QY 360 ISAFV-KTOSAWNS---DSEKFPDHLQKALLYNNNSKLTQSAANSYRILNRTPTNQTG 415
Db 1607 YFTLVKTNLGVVEIFCNKNKIDIFY--GEWKTSSKSTPSGNGRNVVVEPIYNDPTG 1664
QY 416 KK---DPRYTADRTIGGYBFLLANDVNSPVVQARQLNWLHFLMNF--NIYAN 465
Db 1665 EDISTSDFSYEPYCIDRYIN--KVLIAPDLYTS-----LININTNYYSN 1708
QY 466 D-----PDANFDS-----IRDAVDNVDADLLQ---IAG- 491
Db 1709 EYYPEIIVLNPNTFHKKNINLNDSSFBYKWSSTEGSDFILVRYLBSNKKILQKIRKI 1768
QY 492 -----DYLKAAG--IKNDKAAN-----DHL--SLEAWSYNDTPYLH 526
Db 1769 LSNTOSFNKMSIDFDIKLSIGYIMSFPKSPNSNELDRDLHLPKIID---NKTYYD 1824
QY 527 DPG---DNMINNDRLRSLLSLAKPLNQRSGMPLTNSLVNRTDDNAETAAPVPSYF 583
Db 1825 EDSKLVKGLINNN---SLFYDPIEFNLVTGWQTI--NGKKYYFDINT-GAALTSYKI 1877
QY 584 IRA-----HSEVQDLI-----RNIIRTEINPNVGY--SPTTEEIKKAPFI 623
Db 1878 INKGHFYFNNDGVMQGVKFGPDGEYFAPANTQNNIEGQAIYVQSKFLTLNGKKYYFD 1937
QY 624 YNKD-----LLATEKKYTHYNTALSYALL-----LTKXSSVP 655
Db 1938 NNSKAVTGWRIINNEKYFNPNNAAVGLQVIDNKNYFNPDTAIISKGWQTVNGSR-- 1995
QY 656 RVYVYGDMPFDDG-QYMAHTIN---YEAITLLKARIKYVSG-----G 694
Db 1996 ---YYPD--TDTAIFNGYKTIQDKHFYDSDCVKIGVFTSNGPEYFAPANTYNNIEG 2051
QY 695 QAMRNQOV-----GNSIITSVRY--GKGALKATDTGDRTRTTSQVAVIEGNN 740
Db 2052 QAIYVQSKFLTLNGKKYYFDNNSKAVTGWQITDSKKYFNTNTAEEAT---GWQITDGK- 2107
QY 741 PSLRLKASDRVVVMGAHKQAYRPLLLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFT 800
Db 2108 -----KYFNTNTAEEATGWQ-----TIDGKKYFNTNTAIASTGYTIINGKHFF 2153
QY 801 AADIKGYANPQVSGYLGVWVPVGAADQDVRVAASTAPSTDGKSV-HQNAAL----- 851
Db 2154 NTD-----GIMQGVKFGPGGEYFAPANTDANNIEGOAILYQNEFLTNGKKYY 2203
QY 852 ---DSRVMPGFSNFQAFATKEEYV--NVVIAK-----NVDKFAEWGVTDEMAPQVYS 901
Db 2204 FGSDSKAV---TGWRIINNNKYFNPNNAAIAIHLCTINNDKY-----YFS 2246
QY 902 STDGSLDSVIONGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKWADWVPDQM 961
Db 2247 ---YDGILONGY-----ITIERNNFYFDANNESKQWTVGF-KGPNGEYFAP--- 2289
QY 962 YALPEKEVVVTRVVKYTPVAGSQI---KNTLYVVDGKSSGKQDQQAQYGGAFLEELQAKY 1019
Db 2290 -----ANTHNNNIEGOALVYQNKELTLNGKKYYFDNDSKAVTGWQITDGKKY 2336
QY 1020 PELFARKQISTGVMPDPSVKIKOWSA-----KYFNGTNILGRGAGYVLKQDQANTVYFSLVS 1075
Db 2337 YFNLTNAAEATG-----WQITDGKKYFNFNLNTAEEATGWQITD-GKKYYF--N 2381
QY 1076 DNTFLPK---SLAVNPNH---GTSSSVTGLVFDG-KGYVYVYSGNOA-----KNA 1118
```

```
Db 2382 TNTFIASGYTSINGKHFFYFNTDGIQMOIGVFGKPGNGFEYFAPANTDANNIEGOAILYQNK 2441
QY 1119 FISLGNWYVYFDNNGYVMTGAQSIANGANYFSLN-GIOLRNAIYDNGKNVLSYYGNDGRR 1177
Db 2442 FLTLANGKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQITNGKKY--YFNTNTSI 2499
QY 1178 YENGYYLFGQWRVFO-NGIMAVGLTRVHGAVOYE-----DASGFOAKGQFITTADGKL 1231
Db 2500 ASTGYTIISGKHFFYFNTDGIQMOIGVFGKPGDGEYFAPANTDANNIEGOA---IR 2550
QY 1232 YFDRDSGNQISNRFRVNRNSKGEWFLFDHNGVAVTGTVTENGQRLYEKP----- 1278
Db 2551 Y-----QNRFLYLHDNIYY-FGNNSKATGWITIDGNRYEYFEPNTAMGANGYKTI 2600
QY 1279 -----NGVQAKGEFIRDANGYLRYYDPNS-----GNEVR--NRFRVNRNSKGEWFLPD 1322
Db 2601 NKNFYFRNGLPQIGVF-KGSNGF-EYFAPANTDANNIEGOAIRYQNRFL-HLLGKIYYFG 2657
QY 1323 HNGIAVTCARVUNG 1336
Db 2658 NNSKAVTGWQITNG 2671
```

RESULT 2

```
US-11-052-554A-217
; Sequence 217, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 217
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-217
```

```
Query Match 4.3%; Score 307; DB 7; Length 726;
Best Local Similarity 27.4%; Pred. No. 4,5e-11;
Matches 97; Conservative 42; Mismatches 97; Indels 118; Gaps 14;
```

```
QY 983 AGSQ-IKN-TLYVVDGK--SSGKQQAQYKGAFLBELOAKYFELPARKQISTGVMPDPSV 1038
Db 52 ASSQKVENOTSNOVEAKTDSANKDPQKGTG-----SVATDAPSMNSA 93
QY 1039 KIKQWSAKYFNGTNILGRGAGYVLKQDQANTVYFSLVSDNTPLPKSLVNPNHGTSSTVGL 1098
Db 94 NNMGSQDKQ-NTVNBISSDSQTKTDEQTD-----LPQNSFKQOS----- 132
QY 1099 VFDGKGYVYVYSTSGNOAK---NAFISLQN-NWYFYDNNGYVMTGAQSIANGANYFSLNSG 1153
Db 133 -----AHVMTTEAEKTPSHSINTFVNDGNGWYVLGADGRNVTSHTIGGKTMVYF--- 183
QY 1154 IQLRNAIYDNGKNVLSYYGNDGRRYENGYYLFGQQWRYFQNGIMAVGLTRVHGAVOYFDA 1213
Db 184 -----AQDGK----- 188
QY 1214 SGFOAKGQFITTADGSKLYBFDSDSGNOISNRFRVNRNSKGEWFLPDHNGVAVTGTVTENGOR 1273
Db 189 ---QYKGAFAQDSDGNKHYYDRDSGEMWMTNRFV-NDQGNWYVYLNNDGVPVTSITVNGOS 244
QY 1274 LYFKPENGVOAKGEFIRDANGYLRYYDPSNGNEVRNRFRVNRNSKGEWFLPDHNGIA 1327
```



```
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 42
; LENGTH: 2367
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-42

Query Match          3.6%; Score 255.5; DB 7; Length 2367;
Best Local Similarity 18.1%; Pred. No. 3.2e-07;
Matches 289; Conservative 240; Mismatches 565; Indels 503; Gaps 74;

QY 34 SLVKAADSDRQQAQVTEQASLVTTSEAAKELTATDTSTATSQPTATVTDNVSTTN 93
DB 768 SLIK-DISSEKVISFNPKENKIIVKSNLPKELSTILLQEIIRNNSSD--LELEKVMLE 824
QY 94 QSTNTANTANFVVPKPTTSEQAKTNDSDKIITTSKAVNRLTATGKFPVANNNTAHPKTV 153
DB 825 CEINVISIETQVBERI--EBAKS-----LTSDSINYIKNEFKLI-----ESI 866
QY 154 TDKIVPIKPKIKQPSLSODDAAALG--NVKNIRKVGKYYVYKEDGTLOKNYALN 210
DB 867 SEALCDLQKQ--NELSDSHIFSEDISETDEGSIRFINKETGESIFVETEKTIPISEYANH 925
QY 211 INKG-----TFEFDGTGALSNNLTPSKKGNL--TNNDNT-----NSFAQYN----- 249
DB 926 ITEBISKIKGTIFDVGKL-----VKKVNLDTTHEVNTLNAAFIQLSIEYNSKESL 979
QY 250 -----QVYS-----TDVANF-----EHDVHYLTASSWRPKYILKDG 281
DB 980 SNLSVAMKVQVYAQLFSTGLANTITDAAKVVVELVSTALDETIDLLPTLSEGLPIATIIDG 1039
QY 282 -----KTWTOSTEKDFRPLMTWPPQETQRY-----VNYMNA-----QLGIHQTY 323
DB 1040 VSLGAQIKELSETSD-----PLLR-----QEIEAKIGIMAVNLTTATTAILTSSLSGIASGF 1090
QY 324 NTATSPLOQLNLA--QTIQTIEBKITAENKTNMLRQTIISAFVKTQSAMNSDSKPF-- 378
DB 1091 SILLVPL-AGISAGIPSLVNNELVLRDKATKVVDYFKHV--SLVETEGVFTLLDDKVMWQ 1147
QY 379 DDLQKQALLYSNNKSLTSQANSNYRILNRTPTNQTGKDPKRYTADRTIGGVFEFLANDV 438
DB 1148 QDQLVISEIDFNNSNIVLGKCE-----IWRMEGGSGHTVTDDI 1185
QY 439 DNSNFPVQAEQLNLWELHLMFNGNIYANDPDANFDSIRVDADVNDVADLLOIAGDYLLKAAK 498
DB 1186 D-----HF-FSAPSITTYREPHUSI-----YDVLVQKEELDLSK 1218
QY 499 GIHKNDKAANDHLSILEASVNDTP---YLHPDGDGNMNM----- 535
DB 1219 DLAVLPNAPNR---VFARWETGWTGLRSLENDGTCKLLDRIRDNYEGEYFWYFAPADA 1274
QY 536 -----DNRLRLSLYLAKPLNORSGNWNPLITSLNRT-----DDNAETAAPVS 580
DB 1275 LITTLKPRYEDTNIRNL-----DSNTRSFIVPIITTEYIREKLSYFSGSGGYALPL 1328
QY 581 YSF-----TRAHDSV-----ODLIENIIRT----- 601
DB 1329 SQYNGINIESESVDIIVDNNVVDVTIESDKTKKGLBIEGLISTLSIENKILNLSH 1388
QY 602 -----EINPNVVGVSFTTEBKAFPIYNKDLLATEKKYTHNTALSYALLTNKSSVP 655
DB 1389 EINFSGVNGSGFVSLTFSILEGINAIEVDLSK-----SYKLLISGELKIL 1437
```

```
QY 656 RV-----YYGDMETDDGOYMAHKTINVEALETLLKARIKYVSGS--QAMENQQVQ 703
DB 1438 MLNSNHIOOKIDYIG--FNSELQ-----KNIPYFVDSSEKEN-GFINGSTKEGLFVSELP 1490
QY 704 NSEIITSV-----RYGKALKATD--TCDRTTRTSGVAVIERGNPNPSLRKLASDRVV 752
DB 1491 DVLTLISKYVMDSDKSPFCGYSSNNLKDKVKVITKDNVNILTYGVLKDDIKISLSLTQDEKT 1550
QY 753 VNMGAHKNQAYRPLLLTTDNGIKAYHSDQBAAGLVVRVTDNRGELI--FTAADIKY--AN 809
DB 1551 IKLNSVHLDSEGVABEILKFMNRKGTNTSDSLMSPLESMTNKSIFVNFLOSNKIFILN 1610
QY 810 PQVSG-----YLGWVVPVGAADODVRAAST 836
DB 1611 FIISGTTSIGOFEPCDENNNIOPYFIKNTLETNYTLVGNRQNMIVEPNYDLDSDGDI 1670
QY 837 APS2TDGKSVHQNAALDSRV-----MFEQFSNFQAPATKEEYTNVVI 878
DB 1671 SSTVINFSQKLYLGIDSCNVKVIISPNITDEINITPYVETNNTYPEVIVLDANYINEKI 1730
QY 879 AKNVDKFAEWGVTFEMAPQYVSS2DGS-----FLDSVION- 914
DB 1731 NVNIN-----DLSIRVWMSNDGNDFILMSTSEENKVSQVKIRFVNVFKDKTLANK 1780
QY 915 -GYAFTDRYDLGISK-----PNKYGTADDLV--KAIKALHSK-----GIKWA-DW 956
DB 1781 LSFNFSQKQDVPVSEIILSFTPSYV--EDGLIGYDLGLVSLYNEKFYINNFGMVSGLIY 1838
QY 957 VPDQWYAL--PEKEVVT--ATRVDK-YGTPVAGSQIKNTLYVVDGKSSGKQQAQYKGA 1010
DB 1839 INDSLYYKFPVNNLITGVTVDGDKYFNFINGGAASIGETIIDDKNYFNO----- 1891
QY 1011 FLEELQAKYPELFAKQISTGVPMDPSVKIKQMSAKYFNGTNILGRG-----AGYVL 1062
DB 1892 -----SGV-LQTGVFSTEDGKFYEPANTLDLENLEGEAIDFTGKLI 1931
QY 1063 KQOATNTYFSLVSDN--TFPLPKSLVNPVNHGTSSTVGLVFDGKYV---YYS2SGNQA 1115
DB 1932 IDE--NIYY--FEDNYRGAWEKELDEGMHFSPB-TGKAFKGLNQIGDDKYFYNSDGMV 1986
QY 1116 KNAPISLGNWYFDFNNGYVMTGAQISNGANYFYLSNGIQLRNIAY--DNGNKVLSSYGN 1173
DB 1987 QKGFVSINDNKHYPDDSGVMKVGYTEIDGKHFFAENG-EMQIGVFNTEDGKFYFAHNE 2045
QY 1174 DGRYENGYLFGQOMRYFQNGIMAVGLTRVHGAQYVFDASQFQAKGOFITTADGKLRYF 1233
DB 2046 DLGNEE-----GGEISY-----SGILFNFNKIIYFDDSDS-FTAVVGWKOLEDGSKYYF 2091
QY 1234 DRDSNQISNRFVRNSKGEWFLFDHNGVAVTGTVTTFNGQRLYFKPNGVQAKGEFIRDANG 1293
DB 2092 DEDTAAEYIGLSLIND-QQYV-FNDDGIMQVGFVTINDKVFYFSDGIIIESG----- 2141
QY 1294 YLRYVDPNSGNEVRNRFVRNSKGEWFLFDHNGIAVGTG 1330
DB 2142 -----VQNIIDNNFYIDDNGIIVQIG 2161
```

RESULT 5

```
US-11-013-759-11
; Sequence 11, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loomis, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:Jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
```

```
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
Us-11-013-759-11

Query Match      3.5%; Score 251.5; DB 7; Length 2314;
Best Local Similarity 18.6%; Pred. No. 5.4e-07;
Matches 293; Conservative 194; Mismatches 576; Indels 509; Gaps 67;

QY 18 TVSIASAVVTLISLSGLVKAUSTDDRQQAUTESQASLVT-----TS 59
DB 776 SVGFVSTYNTVDFIDGNATKATKYDE-----TNQTSKVTYDVNVDEKTIELTDGNGKTN 830
QY 60 EAAKETLTATDTSTATSATSQPTA-----TVTDNVSTTNSQSTNTTANTANFVVKPT 110
DB 831 KIGVKITTLTTNANGKATNFSTTNDALVNAKDIAENLNTLAKEIHTTKGTADTALQTF 890
QY 111 TTSEQAKTDNSDKIITTSKAVNRLTATGKFVPANNNTAHPKTVTDKIVPIKPKIGLKQP 170
DB 891 KVKDKGATD--DETIVGK-----DGTQNGKTV-----916
QY 171 SLSQDDIAALGNVNRKVNKGYKYYKEDGTLQKNYALNNGKTFPFDETCAL--SNNT 228
DB 917 -----NTLKLKGENGLTVATNKDGTV-----TFGINTQSLKAGDST 953
QY 229 LPSKKGNTNNDNTNSFPAQYQVSTVDANFEHVDHLYTAESWYRPKYILKD--GKTWTQ 286
DB 954 TLNKDGLSIKNPASNEQIQ-----VGADGVKFAKVDKGNSSGTIDGTSRITKQDQIGFTGAN 1009
QY 287 STEKDFRPLMTWPDQ-----ETQROYVN-----YMAQLG--IHQTYNTATS-----328
DB 1010 GSLDTTKPHLT---KDKLKVEVEITNTGINAGGKIITNIQSGDITQNSDNDVTCGRVYD 1066
QY 329 -----PLOLNLAQTIQKIEKITAENWNWLRQTSFAFKVTSQAMNSDSKPPDDHLQ 383
DB 1067 LKTELESKINSAAKTAQNSLHEFSVADEQGN-----HFTVSNPYSSYDTSKTSDD---1115
QY 384 KGALLYSNNSKLTQANSNYRLNRPNTNOT--GKQDPRYADRTTGGVEFFLLANDVDNSN 442
DB 1116 --VITFAGENGITTKVNGK--VVRVGIDQTKGLTPKLTGVNNGN--KGIVIDSQKQGN 1168
QY 443 PVVQAEQLNLHLFLNFGNFIYAN--DPDANFDSIRVDVNDVADLLQIAGDYLKAAKGTH 501
DB 1169 TIT-----GLSNTLANVTNDGAGHLSQGLANDTDKTRAASIGDVNLNAGFNQ 1216
QY 502 KNDKAANDHLSLEAWSYND-----TPYLHDDGDN-----INMDNRLRLSLLYSLAKP 550
DB 1217 GNGEAV--DFVSTYDVFDFIDGNATKATKYDYDDTSKTSKVYDVNDVND-----KT 1264
QY 551 LNQRSGMPLTNSLVNRTDDNAET---AAPSYSFIRAHSEVQDLIRNIRTBINPV 607
DB 1265 IEVTSDDKGLVKVTKTLTKTSANGNATKFSAADGDALVKASD-----IATHLN--TL 1313
QY 608 VGYSFTTEEIKKAPIYNKDLATLATEKKYTHYNTALSAYALLTNK---SSVPRVYVYDMFT 664
DB 1314 AGD-----IQTKAGASQASSASYVDAGNKVIYDSTDKKYY--QVN 1353
QY 665 DDGQVMAHKTINYEAI-----ETLLKARIKYVSGQAMRNQOVGNSEIITSVRYKG 716
DB 1354 DKGQVDKNEKVAKDKLVAQAQTPDGTLAQMVKSV-----INKEQVNDANKKQGINEDNA 1408
QY 717 ALKATDGDTRTSGVAV-----IEGNPNSURLKASDRVNVNMGAHKNQ 762
DB 1409 FIKLENAAKDKTKNAAVTVVDGLNVAQAQTPITFAGDTGTTAKKLGETLTIKGGQTDN 1468
QY 763 AYRPLLTTDNGIKAYHSDQEAAGLVRYT-----NDREGELIFTAADIKGVAN 809
DB 1469 -----LTDNNIGV-----VAGTDGFTVKLAKDLTNLNSVNAAG-----TRIDEKGISF 1511
QY 810 PQVSGYLGVVFPVGAADQDVRVAASTAPSTDGKSVHQNAALDSRVMEFGFSNFQAPATK 869

RESULT 6
US-11-013-759-3
; Sequence 3, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaaki, Ken
; APPLICANT: Yang, Yan ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:Jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-3
Query Match      3.4%; Score 245.5; DB 7; Length 1992;
```

Best Local Similarity 19.9%; Pred. No. 9.9e-07;
Matches 304; Conservative 190; Mismatches 615; Indels 421; Gaps 66;

QY	10	RKVKKRWVTVSIAASVVVTLTSLGSSLVKADSTDRQAQTSESQASLVTTSEAAKETLTAT	69
Db	458	KQLKVGSVAITDINGIDAGNKKI SNLAGSSAND--AVTIEQL-----KAAPTLNA- 507	
QY	70	DTSTATSATSQPTATVDNTYSTNQSTNTTANTANFVVKPPTTSEQAKTDNSDKIIITSK	129
Db	508	--GAGISVT--PT-----EISVDAKSGNVTAPTYNIGVK--TTELNSDGTSGFSVKGS	555
QY	130	AVNLRTATGKFVPANNNTAHPKTWTDKI--VPIKPICKLQPSLSQDDIALG--NVK	185
Db	556	GTNLSLVTAHLASLYLNEVN-RTADSALQSFTVKEEDDDANAIVAKO TTKNAGAVSIL 614	
QY	186	NIRKVGKYYYYKEDGT---LOKNYALNINGKTFEPDETGCALSNN TLP SKGN-----	235
Db	615	KLKGKGLTVATKKDGTVTGLSDSGLTIGKST-----LNNDGLTVKD TNEQIOVG 666	
QY	236	----ITNNDNTNSFAQYNQVISTVANFEHVDHYLTAESWYRPK--YILLDKGTWTQST	288
Db	667	ANGIKPTNVNGSNP-----GTGIANTARIIT-----RDKIGFAGSDGAV-----	704
QY	289	EKDFRLMTTWPDQETQRQYVNMMAQLGHOTYN-----TATSPLQLNLAQT	338
Db	705	-----DTNKPYLDQKLOGVNVKITNTGINAGGKAITGLSPTLPSIADQS	749
QY	339	IQTKEEKITAENKTNWLROTI S AFVKVTSQAMNSDEKPPDDHLQKCALLYSNS-----	393
Db	750	-SRNIE-----LGNTIQDKSNSAA-----SINDILNTGFNLKNNNPIDFV	790
QY	394	---KLTSQANSRYILNRTPNTGKDPRTADRTIGGYEFFLLANDVDNSNPVQAEOL	450
Db	791	STDVIDVFANGN--ATTATVHTDTANKTSKVVDNVVD TTIH LGTGDDNKLGVKTKL	848
QY	451	NWL-----HFLMNFIGNIYANDPDANFDSIRVDAVDNV DADLLQ IAGDYLKAAKGHIKNDKA	506
Db	849	NKTSANGNTATNF-NVNSSDEDA-----LVNAKD-----IAENLNTLAKEIHHTKGT	894
QY	507	ANDHLSILEAWSYNDPFLYHDDGDMMINMDNRLLSLLYLAKPLNQSRGMNPLI---	563
Db	895	ADTALDFTVKKVDENN--NADDANAIIVGOK-----NANNQVNTLT L KGXN	939
QY	564	SLVNRD DDNAE-TAAVPSYSFI AHDSEVQD---LI RNIIRTE-----INPN 606	
Db	940	GLNIKTDKNGTVTFGINTTSGLKAGKSTLNDGGLSIKNPTGSEQIOVGADGVKFAKVNN	999
QY	607	VW---GVSFTTBEBIKAFEIYNKDLLATEKKYTHYNTALSYALLLT WNKS VPRYYGDMF	663
Db	1000	GUVGAGIDGTR-----ITRDEIGFGTGNGSLD-----KSKPHL-----	1033
QY	664	TDDGOYMAHKTYNEYALETL L KARIKYVSGQAMRNQQVGNSEII TSVR KGKALKATDT	723
Db	1034	SKDGINAGGKKITNIOSGEIAQNASHDVA CGKTYDLK TELENISSTAKTAQNSLHFBSV	1093
QY	724	GDR-----TTRTSGVAVIEGNPNPSLR LKASDRVVNVN GAAHKQAYRPLL	768
Db	1094	ADRGQNNFTVSNPYSSYDTSKTS SDVIT FAGEN-GITTKV-NKGVVRVIGIO T KLGITPKL	1151
QY	769	LTTDNGIKAYHSQOE-----AAGLVRYTNDRGELIF TAAD IKGANPQVSGYLGVW	819
Db	1152	TUVNNGKGVIDSQNCQNTITGLSNTLANVTNDKGSVRTTEQG-----	1195
QY	820	VPVGAADQDVRAASTAPSTDCKSVHQNALDSRVMPFGSFNFQ AFATK KEEYTNVIA	879
Db	1196	---NIINKDEKTRAASIVD-----VL SAGF-NLQ G--NGEAVDFVST	1231
QY	880	KNVDKFAEMGVTFENAPQVVSSTDCGSFLDSVIONGYAFTDRYDLGI-----	930
Db	1232	YDTVNFADGNATTAKVYTDTSKTSKVVDNVVD TTIEVKDKLGVK TTTTLSTGTGAN	1291
QY	931	KY-----GTADDLVKA--IKA-----LHSGIKIKYMWADVWD	959

```

1292 1292 KPALSNQATGDALVKASDIVAHLNLTSGDIQTAKGSAQNNASAGYVDADGNKIYDSTDN 1351
Qy 960 QMYALPE-----KEYVTATRYDKYGTGP---VAGSQIKNTLYVVDGKSGKQDQAQYKG 1009
1352 1352 KYIQAKNDGTVDKTEKAVAKDKLVAQAQTPDGTTLAQMMNVKSVINKEQVNDANKKQGINEDN 1411
Qy 1010 AFLEELOAQYPELFAKQISTGVPMQPSVKIKQWSAKYFNGTNIILGRGAGYVLKQATYNT 1069
1412 1412 AFVKGLEKAAASD---NKTKNAAVTVGDLNAVAQTPLTTFAGDGTGTTAKKLGSETLTIKGGQT 1468
Qy 1070 YFSLVSDNTE-----LPKSLVNPENHGTSSSVTCLVFDGKGYYVYVSTSGNQAKN 1117
1469 1469 DTWKLTNNIGVAVAGTDGTVFKLAKULTNLN---SVNAGGTKIDDKGVSVFVDSGGQAKAN 1525
Qy 1118 AFISLGNMYTYFDNNGMYMT---GAQSGINGANYFYLS---NGIQLRNAIYDN--GNKV- 1167
1526 1526 TPVLSANG---LDLGGKVISNVGKGTQDTPAANVQQLNEVRNLLGLGNAGNDNADGNQVN 1582
Qy 1168 ---LSYYGNDGRR---YENGYYLFGQ---QWRYFQNGIMAVGLTRFVHGAVQYFDASGF 1216
1583 1583 IADIKKDPNSGSSSNRTVIKAGTVLGGKGNNDTEKLTAGGIQGVGDK----- 1629
Qy 1217 QAKGQFTTADGKLYFRDRDSGNQISNRFVRNSKGWFLFDHNGVAVTGVTGTFNGQRLYF 1276
1630 1630 -----DGNAGDLSNVVKTQK-----DGSKKALLATYNAAQTNYL 1666
Qy 1277 KPNGVQAKGEFIRDANGLYRYDPSNGNEV-----RNRFVRNSKGWFLFDHNGIAVTGA 1331
1667 1667 TNPAAERALDRINEQ---IRFFHVNDGNQEPVVQGRNGIDSSASGK-----HSAVIGFQ 1717
Qy 1332 RVVNGHASILSLMVFLRESLSQSVKVVSN 1361
1718 1718 AKADGEAAV---AIGRQTOAGNQSIAIGDN 1744

RESULT 7
US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1992
; TYPE: PRP
; ORGANISM: Moraxella catarrhalis
US-11-013-759-13

```

1

Db 556 GTNNSLVTAHLASYLNEVN-RTADSALQSFTVKEEDDDANAITVAKDITTKNAGAVSIL 614
Qy 186 NIKRVNGKYYYKEDGT-----LQKNYALNINGKTFEFDGALSNNITLPSKGN----- 235
Db 615 KLKGNGLTVATKDGTVTFGLSQDSGLTIGKST-----LNNDLGLTVKDTNEIQVIG 666
Qy 236 -----ITNNDNTNSPAQYNQVYSTDVANFEHVDHYLTAEWSRPK--YILKDGKTWTQST 288
Db 667 ANGKFTNVGNSP-----GTGIANFARIT-----RDKIGFAGSDGAV----- 704
Qy 289 EKDFRPLMTWPDQETORQYVYNNMAQLGIHQTVN-----TATSPQLMLAAQT 338
Db 705 -----DTNKPYLDDQKLQVGNVKITWGINAGGKAITGLSPLPSIADQS 749
Qy 339 IOTKIEKITAEKNTNWLQTTISAFVKTSAMNSDSEKPFDDHLOKGLLYNNNS----- 393
Db 750 -SRNIE-----LGNTIQDKXSNA-----SINDILNTGFLKNNNPIDPV 790
Qy 394 ---KLTQANSNYRILNRTPTNQTGKDPRTYADRTIGYEFLLANDVDNSPVVQAEOL 450
Db 791 STYDIDVFANGN--ATTATVTHDHTANKTSKVYDVNVDDTTIHLTGTDNDKGLGVKTTKL 848
Qy 451 NWL-----HFLMNFNIYANDPANFDSIRVDVANDVDLLOIAGDYLKAAKGIHKNDKA 506
Db 849 NKTSAWGTATNF--NNGSDEDA-----LVNAKD-----IAENLNTLAKEIHTTKGT 894
Qy 507 ANDHLSILEAMSYNDTPYLHDDGDNIMNDNRLRLSLLSLAKPLNQSRGMNPLI---TN 563
Db 895 ADTALQTFVKKVDENN--NADDANAITVGOK-----NANNQVNTLILGEN 939
Qy 564 SLVNRDTONAE--TAAPVPSYFTRAHDSEVD--LIRNIIRTE-----INPN 606
Db 940 GLNIKTDXGTVTFGINTTSLGKAGKSTLNDGGLSIKNTPTGSEQIQVGADGVKFAKAVNN 999
Qy 607 VV---GYSFTTEEIKAFIEYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVYGDMP 663
Db 1000 GVWGAIGDGTTR-----ITRDEIGFTGNGSLD-----KSKPHL----- 1033
Qy 664 TDDQYMAHKTINYEAIETLAKIRIKYVSGQAMRNQOVGSEIITSVRYGKALKATDT 723
Db 1034 SKOGINAGGKKTINTQSGEIAQNSHDVATGGKIYDLKTELENKISSHTAKTAQNSLHFSV 1093
Qy 724 GDR-----TTRTSGVAVIEGNNPSLRKASDRVYVNMGAHKQAYRPLL 768
Db 1094 ADEQGNFTVSNPYSSYTSKTSVDITFAGEN--GITTKV--NKGVVVRVIGIDQTKGLTTPKL 1151
Qy 769 LTTDNGIKAYHSDQE-----AAGLVRYNDRGELIFTAADIKGVANPOVSGYLGW 819
Db 1152 TVGNNGKGIVIDSQNGQNTITGLSNTLANVTNDRKGSVRTTEQG----- 1195
Qy 820 VPVGAADQDVRVAASTAPSTDGKSVHQNAALDSRVMEFGFSNFOAFATKKEEYTNVIVA 879
Db 1196 ---NIKDEKTRASIYD-----VLSAGP--NLQG-----NGEAVDFVST 1231
Qy 880 KNVDKFAEWGVTDFEMAPQYVSSSTDGSLDSVIONGYAFTDRYDIGI-----SKPN 930
Db 1232 YDTVNFADGNATTAQVYDDTSKTSKVYDVNVDDTTIEVKDKLGVKTTTLTSTGTGAN 1291
Qy 931 KY-----GTADDLVKA--IKA-----LHSGIKGIKWADWVPD 959
Db 1292 KPALSNQATGALVKASDIVAHLNLTSGDIQTAKGASQANNSAGYVDADGNKVIYDSTDN 1351
Qy 960 QMYALPE-----KEVVTATRVDKYGP--VAGSQJKNLTLYYVDGKSSGDKQQAQYGG 1009
Db 1352 KYQAQKNDGTVDKTEVAADKLVQAQTPDGTILAQMNVKSVINKEQVNDANKKQGINEDN 1411
Qy 1010 AFLELOQAYPELFARKQISTGVPMDPSPVKIKQMSAKYFNGTNIILGRGAGVYLKDAQNT 1069
Db 1412 AFVKGLEKAAAD--NKTKNAAVTVGDLNAVAQTEPLTFAGDTGTAKKLGTELTIKGGQT 1468
Qy 1070 YFSLVSDNTP-----LPKSLVNPNGHTSSSVTGLVFDGKGVYVYSTSGNQAKN 1117
Db 1469 DTNKLTDNNIGVAGTDTGTVKLAKDLTNLN---SVNAGGTKIDDKGVSFVDSQAKAN 1525

RESULT 8

US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loesmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PRF
; ORGANISM: Moraxella catarrhalis
US-11-013-759-4

Query Match 3.4%; Score 245.5; DB 7; Length 2047;
Best Local Similarity 19.9%; Pred. No. 1e-06;
Matches 304; Conservative 190; Mismatches 615; Indels 421; Gaps 66;

Qy 10 RKVKRWTVTSIASAVVLTLSLSGSLVKADSTDDRQQAVTESQASLVTTSSEAKETLTAT 69
Db 513 KQLKGVSAITIDNGIDAGNKISNLAKGSAND---AVTIEQL-----KAARKPTLNA- 562
Qy 70 DTSTATSATSOPTATVTDNVSTTNTTANTANFVVKPTTTSSEAKETDSDKIITTSK 129
Db 563 --GAGISVT--PT-----EISVDKSGNVTAFTYNIQV--TTTSLNSDGTSDKESVKGS 610
Qy 130 AVNRLTATGKVPANNNTAHPTVTVDKI--VPDKPKIGKLKOPSLSQDDIAALG--NVK.185
Db 611 GTNNSLVTAHLASYLNEVN-RTADSALQSFTVKEEDDDANAITVAKDITTKNAGAVSIL 669
Qy 186 NIKRVNGKYYYKEDGT-----LQKNYALNINGKTFEFDGALSNNITLPSKGN----- 235
Db 670 KLKGNGLTVATKDGTVTFGLSQDSGLTIGKST-----LNNDLGLTVKDTNEIQVIG 721
Qy 236 -----ITNNDNTNSPAQYNQVYSTDVANFEHVDHYLTAEWSRPK--YILKDGKTWTQST 288
Db 722 ANGKFTNVGNSP-----GTGIANFARIT-----RDKIGFAGSDGAV----- 759
Qy 289 EKDFRPLMTWPDQETORQYVYNNMAQLGIHQTVN-----TATSPQLMLAAQT 338
Db 760 -----DTNKPYLDDQKLQVGNVKITNTGINAGGKAITGLSPLPSIADQS 804

QY 339 IOTKIEEKITAEKNTNWLROTIISAPVKTQSAWNSDEKPFDDHLOKQALLYSNNS----- 393
Db 805 -SRNIE-----LGNTIQDKXGNA-----SINDILTNFNLKNNNPIDFV 845
QY 394 ---KLTSAANSYRILNRTPTNOTCKDPRYTDRTIGGYEFLLANDVNSNPVVOAEOQL 450
Db 846 STYDVIDFANGN--ATTATVTHDTANKTSKVYDVNVDOTTHLGTDDNKKLGKVTYKL 903
QY 451 NWL-----HFLMNFNIYANDPANFDSIRDAVDNVDADLLQIAGDYILKAAGHKNDKA 506
Db 904 NKTSANGNTATNF-NVNSSDEDA-----LVNAKO-----IAENLNTLAKEIHTTKGT 949
QY 507 ANDHLSILEAWSYNDTPYLDHGDNMNMNRRLRSLLSYSLAKPLNQSRGWNPLI---TN 563
Db 950 ADTALQTFVKKVDENN--NADDANAITVGOK-----NANNQVNTLTKGEN 994
QY 564 SLVNRDTONAE--TAAPVPSYFIRAHDSVQD---LIRNIIRTE-----INPN 606
Db 995 GLNKTDRKNGTVTFGINTTSLGKAGKSTLNDGSLSIKNTPTSEQIQVGADGVKFAKVN 1054
QY 607 VV---GYSFTTEEIKKAFIYNKDLLATEKKYTHYNTALSYALLTNKSSVPRVYGDWF 663
Db 1055 GVWGAIDGTR-----ITRDEIGFTGNGSLD-----KSKPHL----- 1088
QY 664 TDDGOYMAHKTINYEAIETLKARIKYVSGGQMRNQOVGNSEIITSVRYKGALKATDT 723
Db 1089 SKDGINAGGKITNTQSGEIAQNSHDAVTGKIYDLKTELENKISSTAKTAQNSLHEFSV 1148
QY 724 GDR-----TTTSGVAVIEGNNPSLRKASDRVVNMGAHKNQAYRPLL 768
Db 1149 ADEQGNNTVNPVSSYDTSKTSVITFAGEN-GITTKV-NKGVVRVGDIDQTKGLTTPLK 1206
QY 769 LTTONGIKAYHSDQ-----AAGLVRYTNRDELIFTAADIKGYANPQVSGYLGW 819
Db 1207 TVGNNGKVIDSQNGQNTITGLSNTLANVNDKGSVRTTEQ----- 1250
QY 820 VPVGAADODVRVAASTAPSDGKSVHQAALDSRVMEGFSNFOATKKEEYNNVIA 879
Db 1251 ---NIKEDKTRAASIVD-----VLSAGF-NLOG-----NGEAVDFVST 1286
QY 880 KNVDKFAEWGVTDFEMAPQYVSSYSDGSLDSVQNGYAFTRYDLGI-----SKPN 930
Db 1287 YDTVNFADGNATTAKVYDVTDSKTSKVYDVNVDOTTHIEVKDKLGKVTTLTSTGTGAN 1346
QY 931 KY-----GTADDLVKA--IKA-----LHSGKIKVMADWVPD 959
Db 1347 KFALSNQATGALVKASDI VAHLNLTLSGDIQTAKGASQANNSAGYVDADGNKVIYDSTDN 1406
QY 960 QMYALPE-----KEVVTATRVDKYGT-----VAGSQIKNTLYVVDGKSSGHKQAKYGG 1009
Db 1407 KYQYAKNDGTVDKTEVAKDKLVAQAQTPDGTGLAQMNKSVINKREQVNDANKKQGINEDN 1466
QY 1010 AFLELOAKYPELPARKQISTGVPMDPSPVKIKOWSAKYFNGTINILGRGAGYVLKDOATNT 1069
Db 1467 AFVKGLEKASD---NKTKNAAVTVDLNAVAQTPLTTFAGDTGTTAKKLGTELTIKGGOT 1523
QY 1070 YFSLVSDNTF-----LPKSLVNPNGHTSSSVTGLVFDGKGVYVYSTSQAKN 1117
Db 1524 DTNKLTDNNIGVAGTGDFTVKLAKDLTNLN---SVNAGGTKIDDKGVSVFVDSQAKAN 1580
QY 1118 AFISLNNWYFDPNNGYMT-----GAQSIGNANYFSL-----NGIOLRNAIYDN--GNKV- 1167
Db 1581 TPVLSANG---IDLGGKVISNVGKGTQDTRAAVQQLNEVRNLLGLGNAGNDNADGNQVN 1637
QY 1168 ---LSYYGNDGR-----YENGYILFGQ---QWRYFQNGIMAVGLTRVHGAQYFDASGF 1216
Db 1638 IADIKKDPNSGSSNRTWIKAGTVJGGKGNNDTEKLATGGIQGVGDK----- 1684
QY 1217 QAKGQFITTADGKLYRFRDSDGNQISNRFVRNSKGWELFDHNGVAVTGTTVTFNGQRLYF 1276
Db 1685 -----DGNANGDLSNVWVKTK-----DGSKKALLATYNAAGQTYNL 1721

QY 1277 KPNGVQAKGEFIRDANGYLRYVYDPSNGEV-----RNRFRVNSKGWELFDHNGIATVGA 1331
Db 1722 TNNPAAIDRINEOG---IRFFHVNDGNQBPVVGQRNGIDSSAGK-----HSAIGFQ 1772
QY 1332 RVVNGHASILSLMVFRLRESSLSQSVKVVSN 1361
Db 1773 AKADGEAAV---AIGRQTOAGNQSIAGDN 1799
RESULT 9
US-11-013-759-7
; Sequence 7, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:JB
; CURRENT APPLICATION NUMBER: US/11/013.759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361.619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-7

Query Match 3.4%; Score 245.5; DB 7; Length 2047;
Best Local Similarity 19.9%; Pred. No. 1e-06;
Matches 304; Conservative 190; Mismatches 615; Indels 421; Gaps 66;

QY 10 RKVKRQWTVSIAVAVTLTSLGSLVKADSTDDDDQQAVTESQASLVTTSAAKETLTAT 69
Db 513 KQLKGVSAITDNGIDAGNKKISNLAGKSAND---AVTIEQL-----KAAKPTLNA- 562
QY 70 DTSTATSATSOTATVTDNVSTTQCTNTTANTANFVVKPTTTSBOAKTDSNDKIITTSK 129
Db 563 --GAGISVT--PT-----EISVDKSGNVTAPTNYGVK---TTNELSDGSDSKFSVKGS 610
QY 130 AVNRLTATGKFPANNTAHPKTVTDKI--VPKPKIGKLPQPSLSQDDIAALG--NVK 185
Db 611 GTNNSLVTAEHLASVINEVN-RTADSALQSFTVKEEDDDANAITVAKDITKNAGAVSIL 669
QY 186 NIKVNGKYYYKEDGT-----LQKYALNNGKTFPFBETGALSNNTLPSKKN----- 235
Db 670 KLKGNGLTVATYKDGTVTFGLSQDSGLTIKST-----LNNDGLTVKDTNEIQVG 721
QY 236 -----ITNNDNTNSPAQNVYSTVDVANFEVDHYLTAESWYRPK--YILKDGKTWQST 288
Db 722 ANGKFTNVGNSNP-----GTGIANTRIT-----RDKIGFAGSDGV----- 759
QY 289 EKDFRPLMTWMPDOETQORQVYNNYNAQLGIHQTYN-----TATSPQLNLAQT 338
Db 760 -----DTNKPVLDDQKLQVGNVKITNTGINAGGKAITGLSPTLPSIADQS 804
QY 339 IOTKIEEKITAEKNTNWLROTIISAPVKTQSAWNSDEKPFDDHLOKQALLYSNNS----- 393
Db 805 -SRNIE-----LGNTIQDKXGNA-----SINDILTNFNLKNNNPIDFV 845
QY 394 ---KLTSAANSYRILNRTPTNOTCKDPRYTDRTIGGYEFLLANDVNSNPVVOAEOQL 450
Db 846 STYDVIDFANGN--ATTATVTHDTANKTSKVYDVNVDOTTHLGTDDNKKLGKVTYKL 903
QY 451 NWL-----HFLMNFNIYANDPANFDSIRDAVDNVDADLLQIAGDYILKAAGHKNDKA 506
Db 904 NKTSANGNTATNF-NVNSSDEDA-----LVNAKO-----IAENLNTLAKEIHTTKGT 949

Db 605 -----WLQNN-GSWYYLNANGDMATGWLQYNGSWYYLNANG 639

Qy 1281 VOAKGEFIRD-----ANGYLR-----YYDNSG-----NEVRNRFVNSKGE 1317

Db 640 DMATG-WVKDGTWYYLEASGAMKASQWPKVSDRWYYNGSGALAVNTTVDGYGVNANGE 698

Qy 1318 W 1318

Db 699 W 699

RESULT 11

US-11-052-554A-229

Sequence 229, Application US/11052554A

Publication No. US20050288866A1

GENERAL INFORMATION:

APPLICANT: Sachdeva, et al.

TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

FILE REFERENCE: 30853/40359A

CURRENT APPLICATION NUMBER: US/11/052,554A

CURRENT FILING DATE: 2005-02-07

PRIOR APPLICATION NUMBER: US 60/589,227

PRIOR FILING DATE: 2004-07-20

PRIOR APPLICATION NUMBER: IN 173/DEL/2004

PRIOR FILING DATE: 2004-02-06

NUMBER OF SEQ ID NOS: 763

SOFTWARE: PatentIn version 3.3

SEQ ID NO 229

LENGTH: 619

TYPE: PRT

ORGANISM: Streptococcus pneumoniae R6

US-11-052-554A-229

Query Match 3.3%; Score 232.5; DB 7; Length 619;

Best Local Similarity 28.2%; Pred. No. 1.1e-06;

Matches 68; Conservative 35; Mismatches 99; Indels 39; Gaps 10;

Qy 1096 TGLVFDGKGYVYSTSGNOAKNAFISLGNWYTFDNGGVMVTGAOSINGANYFISNGIQ 1155

Db 402 TGWQKENGWYFYNTDGSNA-TGWLQNNGSWYYLNSNGAMATGWLQYNGSWYYLNANGAM 460

Qy 1156 LRNAIMDNKNVLSYYGNDGRRYENGYYLFGQWRFYQ-NGTMVGLTRVHGAVOYFDDAS 1214

Db 461 ATGWAKVGS-----WYYLNANGAMATGWLQYNGSWYYLNSNGAMATGWAKVGSWYYLNAN 517

Qy 1215 GFQAKGQFITTADGKLRYFDRDSGNQISNRFVNSKGEWFLFDHNGVAVTGTTFNGQRL 1274

Db 518 G-----AMATGWLQY-----NGSWYYLNANGAMATGWAKVGSWY 552

Qy 1275 YPKPNGVOAKGEFIRDANGYLRYYDPSNGNEVRNRFVNSKGEWFLFDHNGIATVGARVV 1334

Db 553 YLNANGAMATG-WVKDGTW--YYLEASGAMKASQWPKVS-DKYYVY--NGLVYV--GALAV 603

Qy 1335 N 1335

Db 604 N 604

RESULT 12

US-11-052-554A-232

Sequence 232, Application US/11052554A

Publication No. US20050288866A1

GENERAL INFORMATION:

APPLICANT: Sachdeva, et al.

TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

FILE REFERENCE: 30853/40359A

CURRENT APPLICATION NUMBER: US/11/052,554A

CURRENT FILING DATE: 2005-02-07

PRIOR APPLICATION NUMBER: US 60/589,227

PRIOR FILING DATE: 2004-07-20

PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 232

; LENGTH: 690

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae R6

US-11-052-554A-232

Query Match 3.2%; Score 227; DB 7; Length 690;

Best Local Similarity 19.7%; Pred. No. 2.7e-06;

Matches 146; Conservative 110; Mismatches 288; Indels 198; Gaps 31;

Qy 702 VCNSEIITSVRYGKALKATDGTTRTSGVAVTEGNNPSRLRLKASDRVVVNMGAHKN 761

Db 20 VNEPILAAAYVNEPILADTSSSEVIKTKVGSIIQQNNIKYKVLV---VEGNIGTVQVG 76

Qy 762 QAYRPLLLTTDNGIKAYHSDQEAAGLVRYTN-DRGELIFTAAIDIKGYA---NPQVSGYLG 817

Db 77 NGVTPV-----EPEAGQDGKPTTPTKITVGDKVTYTEVASQAFSYYPDETGRV- 126

Qy 818 VWPVVGAAADQDVRVAASTAPST-----DGKSVHQAALDSRVFMFEGFSNFQAFATKBEY 873

Db 127 VYYP-----SSITPSSIKKIQKGFHGSKA--KTIIFDKGSQLEK----- 165

Qy 874 TNVVIKNDVKAENGVTDFEMAPQYVSVSTDGSFLDSVIONGYAFTDRYDL-----GI 926

Db 166 ----IEDRAFDFSELEBIELPASUEYIGTSAPFSQKLKLTFFSSSSKLELISHAPANL 221

Qy 927 SKPNKYGTADDLVKAIKALHSK-----GI-----KVMADWV 957

Db 222 SNLEKL----TLPKSVKTLGSNLFELTTLXKHVDVEEGNESFASVDGVLFSKDKTQLIYY 277

Qy 958 P-----DOMYALPE-----KEVVTATRVDKYGTTPVAGSOIK-----NT 990

Db 278 PSQKNDSEYKTPKETELASYSFNKNVLYLKLELNEGLEKIGTFAPADAIAKLEELSPNS 337

Qy 991 LYVVDKSSGDKQQAQKYGAFLLELQAKYPELPARKQISTGVPMDFSVKIKOWSAKY--F 1048

Db 338 LETI-----ERLAFYGNLELKEILPDNVKNFKGHVWNGLP-----KFLTL 378

Qy 1049 NGTNILRGAGYV-----LKD---QATNTVFSLVSDNTFLPKS----- 1083

Db 379 SGNNINSLPSPFLSGVLDSLAKELHIKNKSTFESVKDTPAIPETVKFYVTSEHIKDLKS 438

Qy 1084 -----LVNPNHGTSSSVTGLVFDGKGYVYSTSGNOAKNAFISL 1122

Db 439 NLSTSNDIIVEKVDNIKQETDVAKPKKNSQGVGVGWKD-KGLWYYLNSGSMATGWVKD 497

Qy 1123 GNNWYTFDNGGVMVTGAOSINGANYFISNGIQLRNAIYDNGKNVLSYYGNDGRRYENG 1182

Db 498 KGLWYYLNSGSMATGWVKDKGLWYYLNSGSMATGWVKDKG---LWYYLNSGSMATGW 554

Qy 1183 YLFGQWQRYF-QNGIMAVGLTRVHGAVOYFASGFGQAKGQFITTADGKLRYFDRDSGNQI 1241

Db 555 VKDKGLWYYLNSGSMATGWVKDKGLWYYLNSGSMATG---WVKDKGLWYYLNSGSG-M 610

Qy 1242 SNRFVNSKGEWFLFDHNGVAVTGTTFNGQRLYFKPNGVOAKGEFIRDANGYLRYYDPN 1301

Db 611 ATGWVKD-KGLWYYLNSGSMATGWVKDKGLWYYLNSGSMATG-WVKVSGKW--YYTVN 666

Qy 1302 SGNEVRNR-----FVRNSKGEW 1318

Db 667 SGDLLVNTTTPDGYRVNANGEW 688

RESULT 13

US-11-052-554A-91

Sequence 91, Application US/11052554A

Publication No. US20050288866A1

GENERAL INFORMATION:

APPLICANT: Sachdeva, et al.

TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE


```
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 260
; LENGTH: 1647
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-260

Query Match      3.1%; Score 223.5; DB 7; Length 1647;
Best Local Similarity 18.4%; Pred. No. 1.6e-05;
Matches 312; Conservative 228; Mismatches 597; Indels 563; Gaps 82;

QY 1 MEKKVRFKLRKVKRWVTVYSIASAVVTLTSLGSLVKADSTDDRQAV-----TESQASIV 56
DB 1 MEKKQRFSLRKVKSGTFLGSLVFLMMT-----TTVADELSTMSEPTITNHTQQAQHL 56
QY 57 TTSEAAKETLTATDSTATSATS-----OPTATVTDNVSTNQTNTT--ANTANFVVKPT 110
DB 57 TTWELSSAESKSQDTSQITPKTNREKEQOGLVSEPTTELADTDAAPMANTGPDATQKS 116
QY 111 TTSEQAKTNSKIITTSKAVNRLFATGKFPVANNHTAHPKTVTKIPIKPKIGKLGKOP 170
DB 117 ASLPLPVNTDHWKVTGAWDKGYKGQKVVAVIDTGDPAHQSMRISDV--STAKVK-- 172
QY 171 SLSLODDIALGNVKNIRKVGKYYKEDGTGLQKNYALNNGTKFFPDETGALESNNITLP 230
DB 173 ---SKEDMLARQAKAGI-----NYGWIINDKVVF----- 198
QY 231 SKKGMITNNDN--TNSFAQYNOVYSTDVANFEHVDHYLTAESWYRPKYILKDKGTWTOST 288
DB 199 -AHNVVENSNDIKENQFEDFE-----DWAENFE-----PDAEA--EPKAIKHKHVIYRPGST 246
QY 289 EKDFRPLMTW-----WPDQETQROYVNMNAQLGIHOT-----YNTATSPLOQ 331
DB 247 QAPKETVIKTEETDGDHSDIDMTQTDHDKYKESH-----GMHVTGIVAGNSKEAAATGERF 301
QY 332 LNLK-----AQIOTK--IEEKITAENKTNWLRQTISAFVKTSQAWNS 372
DB 302 LGIAPEAQVFMFVANDVMSGAESLFIKAIADAVALGAD-----VINLSIGTANGAQL 355
QY 373 DSEKPFDDHLQK-----GALLYSNNSKLTSQ-----ANSNYRIL-----NRTPTN-- 412
DB 356 SGSKPLMEALBKAKKAGVSVVVAAGNERVYVYGDHDDPLAINPDYGLVGSFPTGRTPTVA 415
QY 413 -----QTGK-----KDP-----RYTADRTIG 428
DB 416 AINSKWVIQRLMTVKELNADLNHGKATYSESDVPKNIKDSIGYDKSHQFAYVKESTDA 475
QY 429 GYEFLLANDVNSNPVQAEQLNWLHPLMNFN-----IYANDPDANFOSIRVDAVD 480
DB 476 GYK---AQVDKQKIALIERDPNKTVDKEMIALAKKHGALGVLIFFNNKPGQSNRSMRLTA-N 531
QY 481 NYVDADLLQIAGDYLKAAGKIHKNDAANDHLSILEA-----WSYNDTPYLH 526
DB 532 GMGIPSAFISHEFGKAMSLQNGNGTSLFSDSVSKAPSKQKGNEMNHFNSWGLTSDGYLK 591
QY 527 DP---GDNMNM--DNRLURLSLLSLAKPLNQRGMNPLITNSLVNRTDDNAETAAPVS 580
DB 592 PDITAPGGDIYTYNDNHYGSGTGTSMASP--QIAGAS--LLVKQYLEKTQPN-----LPK 643
QY 581 YSFIRAHDSVDLLRNII-----RTEINPVVGVSTTEEIKKAFEIYNKOLLATEKKYT 636
DB 644 -----EKIADIYKNNLSNAQIHVNPET---KTTTSPROQGAGLLNID----- 683
QY 637 HYNITALSYALLTNKSSVPRVYVG---DMFTDDGQYMAH-----KTINYEY-----IE 681
DB 684 ---GAVTSGLYVTGKNYGSISIGNITDWTFTD--VTVINLSNKDKTLRYDTLLTDHVD 738
```

RESULT 15

```
US-11-052-554A-281
; Sequence 281, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sacdewa, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
```

```
QY 682 -----TLLKARIKVVSGQAMRNQOVNSBIIITSVRVYKKGALKAT-----DTGDRTRPT 730
DB 739 FOKGRFTLTSLRSLKTYOGGE-----VTVFANGKVTVRVTVMDVDSQFTELTLTKM 786
QY 731 SGVAVIEG-----NNPSL-----RLKASDRV----- 751
DB 787 SNGYILEGFVFRDSQDDQLARVNIIPVGFPGQFENLAVAEESIYRLKSQKGTGYFDES 846
QY 752 -----VVNMGAAHKNQAYRPLLLTTDNGIKAYHSDQEAAG--LVRYTNDRGE 796
DB 847 GPKDDIYVGKHTGLVTLG--SETNVSTKTI---SDNGLHTLGTFRKADGKFILEKNAQGN 902
QY 797 LIFTRAA-----DIKGYANQVSYGLGVWVPVGAADQDVRAASTAPSTDCKSVHQA 849
DB 903 PVLAI SPNGDNQDPFAAFKGVFLRYQGLKASVHASDKE-----HKNP 946
QY 850 ALDSRVMEFGSFNFQAFATKKEEYTNVVIKXNV-----DKFAEWGVTDFEMAPQYVVSITDG 905
DB 947 LWSVESFSGKDNFN-----SDIRFASHTLLGTAFSGKSLTGAEPLPGYHYHVS 997
QY 906 SFLDSVIQNGYAFTRYDLGISK--NKYGTADDLVKAIK--ALHSKGIKVMADWPDQY 962
DB 998 YYPDVVGAQRQEMTFDMLDRQKPVLSQATFDPETNRFPKPELKDQK---LAGVRKDSVF 1054
QY 963 ALPEKE--VVTATRVKYGTVPVAGSQIKNTLYV--VDGKSSGKQOQAKYGGAFLEELQAKY 1019
DB 1055 YLERKDNKPYTITINDSY-----KYSVEDNKTFVERQA--DGSFI----- 1093
QY 1020 PELFARKQISTGVPMDPSPVKIKQMSAKYFNGTNIILGRGAGYVLKQATNTYPSL-----V 1074
DB 1094 -----LPJD--KAKLGD---YVMVEDFAGNVAIAKLGDLHPQLTGLKTIPLK 1137
QY 1075 SDNTFLPKSLVNPNGHTSSSVTGLV-----FDGKGYVYVYSTGNGQ 1114
DB 1138 TDGNYQTKETLKDNLQEMTQSDTGLVTNQALAVVHRNQPSQLTKMNQDFFISPNEDGNK 1197
QY 1115 AKNAFISIGNWY-----YFDNNGYMT-----GAQSGINGANYFISNGIQLRN 1158
DB 1198 DFVAFGLKNNVYNDLTVNVYAKDDHQKQTPWSSQAGASASAIESTAWY-----GITARG 1253
QY 1159 AIYDNGN--KVLSSVYVGNDRRYENGYYLFGQWRYFQNGIMAVGLTRVHGAVQYVDASGF 1216
DB 1254 SKWMPEDYQVYVYRDEHGEKQKQYITISVD-----KXPMITQGRFDITNGVDHTPDKT 1309
QY 1217 QAKGQ-----FITDAGKLYFDRDSNQ-----ISNR-----FVRNSKGEWFLFDHNGV 1261
DB 1310 KALGSSGIVREEVFLAKKNG--RKFDVTEGDKDGTIVSDNKNYIPKPNPDGYSYTIKRDGV 1367
QY 1262 AVT-----GTVTFNGQRLYFKPNGVOAKGE-----F 1287
DB 1368 TLDYVYLLVEDRAGNVSPATLR-----DLKAVGKDKAVVNFGLDLPVPEDKQIVNFTYL 1421
QY 1288 IRDANG-----YLRYYDPNSGNEVRNRFVRNSKGEWFLFDHNGIAVTGARVNGHASILSL 1343
DB 1422 VRDADCKEIEYLEYV--NSGNSL-----ILPYKQYTV-----EL 1454
QY 1344 MVFRLESSLOSQVKSVMNTM 1363
DB 1455 LTYDTNAAKLESCKIVSFTL 1474
```


; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 281
; LENGTH: 5291
; TYPE: PRF
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-281

Query Match 3.1%; Score 223; DB 7; Length 5291;

Best Local Similarity 19.0%; Pred. No. 9.2e-05;

Matches 322; Conservative 187; Mismatches 596; Indels 590; Gaps 77;

QY 17 VTVSTASAVVLTSLGSLVKADSTDDROQAVTESQASLVT-----TSEAAK 63
DB 3326 VTVDTPAVISFNTVAG-----DDVINVEHQIAIISGTATGAVAGDRLVVTIAGQ 3377
QY 64 ETLTATDTS-----TATSATSQPTATVTDNVSTTNQSTNTANTANFVVKPTTTSEQA 116
DB 3378 QVVTSTDSAGNSVGVPAISVSLGADGTVTISATIDSAGNSTQTHNVQVNTAAVSLV 3437
QY 117 KTDNSDKIITTSKANVRLT-----ATGKF-----VPA-- 143
DB 3438 STISGDNLINEAAGSALTLSGTGNFATGTVTVLLNGKYSATTQSNQSMVNVPAAD 3497
QY 144 -----NNNTA---HPKTVTDKIVPIPKIGLQKPSLSQDDIA 179
DB 3498 VAALSDGTSYTVSASAQDSAGNSSTQTHNVQVNTAAVSL-----SVSTISGDNLI 3549
QY 180 -----ALGNVVKIRKNGVKKYKEDGTLQKNYALNNGKTFPFBET 221
DB 3550 NAAEAGSALTLSGTGNFATGTVTV-LLNGKY-----SATIQSNQSMVNVPA--ADV 3601
QY 222 GALSNNLTSPKSGNTNDNTNSFAYNQVYSTDA-----NFEHVDHLYFAESWYRP- 274
DB 3602 AALSDDGT--SYTVSASAQDSAGNSATASRSVAVDLTAPVISINTVSTDDRLNAAEQOQPL 3659
QY 275 -----KYLKDGKWTQSTQKDPPLMTW----- 300
DB 3660 TLNGSTSAEVGTVTVTFGGKTYTATVAAN-----GTWALNVPAVDLAALGQGAQTITAS 3714
QY 301 -----PDQETQROYVNVNNAQLGI-----HOTVNTATSPQLNLAAQTIQKIEKIT 348
DB 3715 VNDRAGNPGQATHATVDTVAFTVATVATVAGDDIINNAE-----QLAGQISGTTTAEVG 3769
QY 349 AEKNTNWLRTISAFVKTSQAMN-----SDSEKPF-----DHLQKGA 387
DB 3770 QTVTVTFNGQTSATVSGGNSVFIPAQQFAGLSGYSYTSATVSDQAGNPGSASRGVT 3829
QY 388 L-----YNNKSLTQANSNYRLNRTPTNQTK-----KDPRTAD-RTIGG 429
DB 3830 LMGDVPVTINTFAGDDVVAEHSGLVISGTTTAPVGQTLTLTLNGKTYTTVTQGS 3889
QY 430 YEFLLAN-DV-----DNSPVVOAEQLNLWFLMNFNIVANDPDANFDSIRVDA----- 478
DB 3890 WSYTLGSADVTALADCNAYVNA-----SVSNAIGTSGSN-HTIITVDLSAPANG 3938
QY 479 --VNVVDADLLQIAGDYLLKAAKGIHKNDK-----AANDHLSIL-----EAW 517
DB 3939 INIDSLOADTGLSADFTSVSPVVVNGSLTAALASNETAQISIDGGTTTTLTVGTGW 3998
QY 518 SYNDTPYLHDDGDNMINMDNRLRLSLLSLAKPLNQRSGMPLTNSLVNRTDDNAETAA 577
DB 3999 RYNDSTRITDGT-----NLYQ-----VRVIDAAGNVGA 4026
QY 578 VPSYSFI---RAHDSVODLIRNIIRTEINPNVGVSYFTTEIKKA-----F 621
DB 4027 TDSQNVVITDTPADPAVKATIAISALTDM--GLINTDFVSTDTTLAVSGTIGATLSAGEF 4084
QY 622 EYKNKDLATEKKYTHYNTALSYALLLTNKSVPVRYVYGDVFTDDGQYMAHKTYNEAIE 681

Search completed: February 11, 2006, 20:57:57
Job time : 17.8246 secs

DB 4085 AQISLDGGVTTTLTVVGTSMVA-----DGHLLTDGTTWTVRV 4124
QY 682 TLLKARIKYVSGQAMN-----QOVGNSELIITSVRYGKALKATD--TGDRTTITSG 732
DB 4125 VDLAGNV-----GQTATQNVVVDTTTPAAKSTITIGISDDTGT-SSSDFTTTLTVRG 4179
QY 733 V-AVIEGNNPSRLKASDR---VVVNMGAHKNQAY---RPLLLTTDNGIKAYH-SDQEA 784
DB 4180 VLGAALGANEFQISTDNGATWVTVVADSLNWSYVDR---TLTNGTTTWOVRVVDL 4235
QY 785 AGLVRYTNDRGELIFTAADIKGYANP-----QVSGYLG 817
DB 4236 AGNVGATSSQSALIDT-----VNPQAQVLTIASISTDTGSSATOFITSDTMLTITGSLG 4288
QY 818 VWPVGAADQDVRVAAS--TAPSTDGKSVHONAAALDSRVMPFSGSNFOAFATKCEYTNV 876
DB 4289 AGLASGEVAQISLDSGATWTLTTLTNGT---QWYTDSTRTLTDGSYVYQVRVLDLAGTGP 4345
QY 877 VIANKVDKFAEWGVTDF---EMAPQVSVSTGDSFLDSVIQNGYAFTRDYLGIKPNKYG 933
DB 4346 VVSKTV-----VVDINPTATPIVSYTD-----DVG-----OROG 4376
QY 934 T-----ADDLVKAIKALHSKGIKVMADWPDQMYALPEKVV-----TATRVDK 977
DB 4377 TLSSSQATDDTTPLLVGLSA-----PLASGEVVYLYRNGLLGCAVTMGA 4422
QY 978 YGTPVAGSOIKNTLVVVDGKSGKQQAQYGGAFLEBLOAKYPELFARKQISGTGPMDFPS 1037
DB 4423 LNWYTSDSLVSAGYATYSGRVVDLAGNITSSSDFVLTVDTSIPTTLA--QITSQTRDIT 4480
QY 1038 VKIK-----QWSAKYFNGTINILGRGAGYVLKQDQATNTYFSLVSDNTFLPKSLVNP 1087
DB 4481 PIISGVITAAALASGOYVEWINGKTYTSEPGCAVVVDPAHNTWYVQLPDTDALTVS--- 4536
QY 1088 NHGTSSTVGLVFDGKGYVYVSTGNAQK-----NAFISLGNW-----Y 1127
DB 4537 --ATAYTVTAQVKSAG-----NGNNANISNGTVTVNAAIDYTPWTWTASKTTAWGLTY 4588
QY 1128 YPDNNG-YWVTGAQSI-----NGANYFSLNSGILQRLNAIYD-NGKV 1167
DB 4589 GLDSHGMWTVLANQQVMQSTDP L TWSKTALTLYQSGNNYATSS-----IADYDRNGTGD 4642
QY 1168 LSYGND--GRYENGYYLFGQWRVYFONGI--MAVGLTRVHGAVQYFDASGFOAQGFIT 1225
DB 4643 LPIRDDYGTGYINGFTNGB--GTFSSAIQVTVGTLTWYGSIVAFDKEG----- 4690
QY 1226 ADGKLRYFDRDSGNQISNRFRVNSKGEWFLFDHN-----GVAVTGTVTTFNGQRL 1274
DB 4691 -DGYLDPMWIGDAGGPDSTFLWNNAGT--LVGNSTTSNGSGSATVGGAVTGVLSLN--- 4743
QY 1275 YPKNGVQAKG-----EFIRDANGYLRYVDPNS-----GNEVRNRFVRNSKG--- 1316
DB 4744 ---EGSGVDLNDNRIDLVLQHTNLNMYVTLSSLNQNGNGTFFWVGQNTTNTFLSGAGSAM 4801
QY 1317 -----EWFLFDHNG 1325
DB 4802 SSSVSWTWADFDG 4816

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 / Search time 28.3806 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-35
Perfect score: 7143
Sequence: 1 MEKKVRFKLKVKKKWTVTS.....VKVNSNTMILIPMKFVIM 1375

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7143	100.0	1375	2	JT0345
2	5285	74.0	1475	2	B33135
3	4168.5	58.4	1592	2	A38175
4	3591.5	50.3	1431	2	A45866
5	3216	45.0	1577	2	T30858
6	3159	44.2	1449	2	T30857
7	3146	44.0	1449	2	T30552
8	3070	43.0	1518	2	A44811
9	3044	42.6	1508	2	T31098
10	2899	40.6	1599	2	S22737
11	2893.5	40.5	1365	2	A41483
12	2870.5	40.2	1290	2	JC5473
13	428.5	6.0	2817	2	B97033
14	308.5	4.3	2710	2	A37052
15	292.5	4.1	2364	2	I40884
16	283	4.0	2178	2	S55805
17	274.5	3.8	563	2	A37184
18	268.5	3.8	2366	2	S10317
19	267.5	3.7	1609	2	S25345
20	255.5	3.6	2367	2	S70172
21	242	3.4	648	2	S10869
22	235	3.3	701	2	H98120
23	235	3.3	1463	2	T30290
24	233.5	3.3	1385	2	D89824
25	233.5	3.3	1939	2	D97316
26	233	3.3	1335	2	T30211
27	232.5	3.3	619	2	A37887
28	232.5	3.3	619	2	A41971
29	230	3.2	1315	2	T28679

ALIGNMENTS

RESULT 1

JT0345
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C:Accession: JT0345; C33135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: JT0345; MUID:89137980; PMID:2976010
A:Accession: JT0345
A:Molecule type: DNA
A:Residues: 1-1375 <UED>
A:Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515
A:Experimental source: GS-5
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: C33135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: UNIPARC:UPI000014E25D; GB:M17361
C:Genetics:
C:Gene: gtfC
C:Function:
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucan:
C:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-14/Domain: signal sequence #status predicted <SIG>
F:35-1375/Product: glycosyltransferase #status predicted <MAT>
F:1126-1145/Domain: cpl repeat homology <CP1>
F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 100.0%; Score 7143; DB 2; Length 1375;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKKVRFKLKVKKKWTVTSASAVVTLTSLSGSLVKADSTDDROQAVTESQASLVTTSE 60
DB 1 MEKKVRFKLKVKKKWTVTSASAVVTLTSLSGSLVKADSTDDROQAVTESQASLVTTSE 60
QY 61 AAKETLTATDTSTATSATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTTSEQAKTDN 120
DB 61 AAKETLTATDTSTATSATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTTSEQAKTDN 120
QY 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVIPKPKIGLQKPSLSQDDIAA 180
DB 121 SDKIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVIPKPKIGLQKPSLSQDDIAA 180

probable RTX famil
choline-binding pr
toxin-like outer m
toxin-like outer m
hypothetical prote
cell wall-associat
hypothetical prote
hypothetical prote
adhesin homolog lm
toxin-like outer m
probable adhesin 2
uncharacterized pr
fibrinogen-binding
hypothetical prote
filamentous hemagg


```
QY 831 RVAASTAPSTGKSVHONALDSRVMPFGFSNFOAFATKBEYTNVIAKNVDKEAEKGV 890
Db 802 RVAASDTASTGKSLHQDAAMDSRVMPFGFSNFOAFATKBEYTNVIANNVDFEVSNGI 861
QY 891 TDFEAPQVYSTGDSFLDSVQNGYAFTRDYLIGISKPNTKGTADDLVKAKIKALHSKGI 950
Db 862 TDFEAPQVYSTGDSFLDSVQNGYAFTRDYLIGISKPNTKGTADDLVKAKIKALHSKGI 921
QY 951 KVMADVPDQMYALPEKEVVTATRVVDKYGTPVAGSQIKNTLVYVDGKSGKQQAQYKGA 1010
Db 922 KVMADVPDQMYALPEKEVVTATRVVDKYGTPVAGSQIKNTLVYVDGKSGKQQAQYKGA 981
QY 1011 FLEELQAKYPELFPARKQISTGVMPDPSVKIKOWSAKYFNGTILGRGAGYVLKDAQNTY 1070
Db 982 FLEELQAKYPELFPARKQISTGVMPDPSVKIKOWSAKYFNGTILGRGAGYVLKDAQNTY 1041
QY 1071 FSLVSDNTFLPKSLVNPNGHTSSSTGLVDFGCKGVVYYST-SGNQAKNAFISLGNWYVF 1129
Db 1042 FNASDITLFLPSSLLG-----KWESGRYDQKGYIYNSSATGDOVKASFITEAGNLVYF 1096
QY 1130 DNGYMTVGAQSGINGANYFLSNGIQLRNAIYDNGKVLVSYGNDGRRY--ENGYLFGQ 1187
Db 1097 KRGDGMVTGAQTINGANYFLENGTALRNTIYTDAGQNSHYIYANDGKRYENENGYQQFGN 1156
QY 1188 QWRYFQNGIMAVGLTRVGAQVYFASGFAQKQFITTADGKLYRFDSDSGNQISNRVFR 1247
Db 1157 DWRYFKDGNMAGVLTVDGNGVQYFDQGVQAKDKIIVTRDQKRYFYDQHNGNAVNTPIA 1216
QY 1248 NSKGWELFDHNGVAVTGTVTFNGQRLYFKPQNGVQAKGEFTDANGLYRYDPSNGNEVR 1307
Db 1217 DKTGHVYLGKDGVAQVTAQTVGKQKLYFEANGQQVKGDFVTSDEGKLYFYDVDSDGMWT 1276
QY 1308 NRVFNSKGEWFLFDHNGIANTGARVNG 1336
Db 1277 DTFIEDKAGNWFLYLGKDGAAVTAQTIRG 1305

RESULT 4
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866
J.Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A:Reference number: A45866; MUID:91100958; PMID:2148600
A:Accession: A45866
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 50.3%; Score 3591.5; DB 2; Length 1431;
Best Local Similarity 51.9%; Pred. No. 4e-169;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 MEKKVRFLKRVKRWVTVSIASAVVTL-TSUGSLVKAADS-----TDRQQA 47
Db 1 METKRYKMKHVKYKHWTVAVASGLITLTGTLGSSVSAETBQQTSDSKVVTOKSEDDKAA 60
QY 48 VTESQASLVTTSEA-AKETLTATDTSTATSQTATVTDNVSTNQS-----TNTTAN 101
Db 61 SSSQTDPAKTKQAQTEQTAQSQANVADTSITKEIPSQNITQANSDDKTVNTKSE 120
```

```
QY 102 TANFVVKPTTTSEQATDSDSKIIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIK 161
Db 121 EAQTSEERTKQSEAOITASSQALTOAKA--ELT----- 152
QY 162 PKIGKLGKOPSSLSQD-----DIAALGNVKNIRKVNKKYKYKEDGTLOKNVALNINGKTF 216
Db 153 -----KORQTAAGBNKPNVDLAI PNK---QIDGKYIYGSDQOPKKNFALT VNNKVL 203
QY 217 FFDE-TGALSNNITLPSKKGNIT--NNDNTNSFAQYNQVYSTDVANFEHVDHYLTAESWYR 273
Db 204 YFDKNTGALTDTSOVQFKGLTKLND---YTPHQIVNFENTSLETIDNTVTTADSWYR 259
QY 274 PKYILKQKWTQSTQKDFRLLMTWPDQETQRYVYVMAQA-LGIHQTYNTATSPLOL 332
Db 260 PKDILKNGKWTASSESDRLFLMSWPDQKQTAIYLYNMNQGLGTGENYATDSQESL 319
QY 333 NLAQTIOITKEKITAEKNTNWLROTISAFVKTSQAMNSDSEKPPD---DHLQKGALL 388
Db 320 NLAQTVOVKLETKLSQTOQOMLRDIINSFKTOPNWSQTESDTSAGEKDHLOGGALL 379
QY 389 YSNNSKLTQASNSRYILNRTPTNQTKGKOPRYTADRTIGGYEFLLANDVDNSPNVQAE 448
Db 380 YSNSDK-TAYANSYRLNRTPTSTQTK- PKYPEDNSSGGYDFLLANDIDNSPNVQAE 436
QY 449 QLNMLHFLMNFNGNIYANDPDANFDSIRVDADVNDVADLLOIAGDYLKAAGKHGKDKAAN 508
Db 437 QLNMLHYLMNYGSI VANDPEANFDGVRVDADVNDVNADLLOI ASDYLKAHYGKSEKNAI 496
QY 509 DHLSTLEAWSYNDPTYLHDDGDNMNMNRLRLSLYSIAKPLNO-----RSGMNP 560
Db 497 NHLSTLEAWSNDPOYNKDTGAQPIDNKURLSLLYALTRPLEKDAKNKNEIRGLEPV 556
QY 561 ITNSLVNRTDDNAETAAPVPSFIRAHDSVQDILFIRNIIRTEINPNVVGYSFTTBEIKKA 620
Db 557 ITNSLNNRSEAGKNSERMANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFLDELKQA 616
QY 621 FEIYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVYVYGDFTDDGQYMAHKTINYEAI 680
Db 617 FKIIYEDMRQAKKYTQSNIPAYALMLSNKDSITRLYVYGDYSDGQYMATKSPYDAI 676
QY 681 ETLKARIKYVSGQAMENQOVGNSE-----IITSVRYGKALAKATDGTDRTRTS 731
Db 677 DTLKARIKIYAAGQDMKITTVGDKSHMDNDYTGVLTSVRYGTGANRATDQGSATKIQ 736
QY 732 GVAVTEGNNPSLRKASRVVYVNMGAHKQAQYRPLLTTPDNGIKAYHSDQEAAGLVRYT 791
Db 737 GMAVITSNNPSLKLNDKQNVIVNMGAAHKQOEYRPLLTTKDGLTSYTSDAAKSLYRKT 796
QY 792 NDRGELIFTAADIKGYANPOVSYGLVWVPVGAADQDVRVNAASTAPSTDGKSVHONAL 851
Db 797 NDKGELVFDASDIQGYLNPQVSGYLAVWVPVGAADQDVRVNAASKANATQGVYSSSAL 856
QY 852 DSRVMFEGFSNFOAFATKBEYTNVIAKNVDKPAEWGVTDFEMAPOVYVSSYDGSFLDSV 911
Db 857 DSQLIYEGFSNFQDFVTIKDSYITNKIAQNQLFKSWGVTSEFEMAPOVYVSSDGSFLDSI 916
QY 912 IQNGYAFTRDYLIGISKENKGTADDLVKAKIKALHSKGI KVMADVPDQMYALPEKEVVT 971
Db 917 IQNGYAFEDRYDLAMSKNKYGSSQODMINAVKALHKSIGIQTADVWPDQIYNLPGEVVT 976
QY 972 ATRVDKYGTPVAGSQIKNTLVYVDGKSGKQQAQYKGAFLBELQAKYPELFPARKQISTG 1031
Db 977 ATRVNDYGEYRKDSIEIKNTLYAANTKSNKGQYQAKYGGAFLELAAKYPSIFNRTQISNG 1036
QY 1032 VPMDSVKIKOWSAKYFNGTILGRGAGYVLKDAQNTYVFSLVSDNTFLPKSLVNPNGHT 1091
Db 1037 KKIDSEKITAWKAKYFNGTILGRGAGYVLKDNASDKYFELKGNQTYLPKQNTN----- 1091
QY 1092 SSSVTLGVFDGKGYVYVYSTSGNQAKNAFI-SLGNWYVFDNNGYVMTVGAQSGINGANYFL 1150
Db 1092 KEASTGFVNDGNGMTFYSTSGVQAKNSFVQDAKGNWYVFDNNGHWYVGLQQLNGEVOYEL 1151
```


474	LNNS--ATSHANSDFRLMNRFTTQGTGR--KYHIDRSGGYELLANDIDNSNPAVQAEQ	530
450	LNWLHPFLMNFENIYANDPDAFNSIRVDADVNDVDADLLQIAGDYLKAAGKHKNDAAND	509
531	LNWLHYIMNIGSILGNDFPSANFQGVRIADVNDVDADLLQIADSYFKEKYRVADNEANAIA	590
510	HLSTLEAWSYNDPTYLHDDGDMNIMDNRLRLSLYLSLAKPLNQRSGMNPILITNSLVNRT	569
591	HLSTLEAWSYNDHQYNKDTKGAQLSDINPLRETLITFLRKSNTGRGSLERIVITNSLNNRS	650
570	DDNAETAAVSYSFIRAHDSVEQDLINRIITETINPNVVGVSPTTEIBKAFELYNKDLL	629
651	SEQKHTPRDQYIFVRHADSEVOAVLANIISKQINPKTGDGTFMTDELKQAFELYNADIA	710
630	ATEKCKYTHYNTALSYALLTNKSSPVRYCYDFTDDGQYMAHKTINVEALETLLKARIK	689
711	KADCKYTOYNI PAAYATMLNKDSITRVYIGDLFTDDGQYMAEKSPPYNAIDALLRARIK	770
690	YVSGGQAMRNQOQVGNSEIITSVRYGKGALKATDTGDRTTTSGVAVITEGNPSRLKASD	749
771	YVAGGQDMKYTKUNGYEIMSSVRYGKABEANOQGTAEITRNQGNMLVLTANRPDMKLGAND	830
750	RVVNVNGAAHKQAYREPLLITTDNGIKAYHSDQEA--AGLVRYTNDRGELFTAAADIKGYA	808
831	RLVNVNGAAHKQAYREPLLKSTGLATYILKSDVPAGLVRYTNDQGNLFTTADDIAGHS	890
809	NPOVSGYLGVMVPVGAADQDVRVAASTAPSDTGKSVHQ--NAAALDSRMVPEGFSGNFQAPA	867
891	TVEVSGYLAWVPVGASENQADTKASS--TKKEQVFESSAALDSQVIEGFSNFQDFV	948
868	TKKEEYTNVVIKKNVDPKAFBWGVTDFEMAPQYVSSSTDGSLDSVIQNGYAFPTDRYDLGIS	927
949	KTSQYTNRYVIAQNAKLKFEKWGITSFEPAPQYVSSQDGTFLDSIIENGYAFEDRYDIAMS	1008

Qy	928	APKAIYADVADVDAIKALHSGIKVMDWVFDQUMIALPFEKEVVIATKVDKIFPVAGSQI	981
Db	1009	KNNKYGSLKDLMDALRALHAEGISAIADWVDPDQIYNLPGEKVVTASRTNSYGTPRPNAEI	1068
Qy	988	KNTLYVVDGSSGKDOQAQYGGAFLELOAKYCELEFARKOISTCVPMDPSVKIKOWSAKY	1047
Db	1069	YNSLIYAARTKTGNDFQGGYGGAFDELKAKYFAIFERVOISNGRKULTTNEKITQWSAKY	1128
Qy	1048	FNGTNTILGRGAGVYLKQDQATNTYFSLVSDNTFLPKSLVNPNGHTSSSVTGLVFD- -GKGY	1105
Db	1129	FNGSNTQGTGARVYLDQDNATNQYFSVKAGOTFLPKQM- ------TEITSGSFRVRGDDV	1180
Qy	1106	VYVYSTGNOAKNAFISLG- -NNWYTFDNNNGYMTGAQSGINGANYFYFLSNGIQLRNAIYDNG	1164
Db	1181	QYLSIGCYLAKNTFIQVGANQWYFYDKNGNKNVTGEQVIDCKKYFFLDNGLQLRHVLRRQS	1240
Qy	1165	NKVLSYGNDGRYENGYYLFG- - -QWRYFO- -NGIMAVGLTRVHGAVQYFD- -ASGFOAK	1219
Db	1241	DGHVYYTDPKGVQAFNGFYDFAGPRQDVRYFDGNGQMYRGLHDMYGTTFYFDEKTIQOAK	1300
Qy	1220	GQFIITTAGDKLRYFDRDSGNQISNRFRVNRSKGE- -WFLFDHNGVAVTGTVTFNGQRLYFKP	1278
Db	1301	DKFIRFADGSTRFYFIDPTGNLAVNRFQAPENKAWYLDNSGYAVTGLQTNGKQYFYDN	1360
Qy	1279	NGVQAQGEFIRDANGVLYRY- -DPNSGNEVRNFRVNRSGKWELFDHNGIAVTGARVNVGH	1337
Db	1361	EGRQVKGHFVTINN- -QRIFYLDGSGEIAPSRFV- -TENNKYWYVDGKVLKGAQVINGN	1417

RESULT 6
 T30857
 Glucosyltransferase - Streptococcus salivarius
 C:Species: Streptococcus salivarius
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T30857
 R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
 Infect. Immun. 63, 609-621, 1995
 A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding
 A:Reference number: Z290909; PMID:95122197; PMID:7822030

RESULT 6
T30857
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-
C:Accession: T30857
R:Stmpeon, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 p
A:Reference number: Z20909; MUID:95122197; PMI

A:Accession: T30857
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: UNIPROT:O55264; UNIPARC:UPI00000B166E; EMBL:L35495; NID:g662378; PID
C:Genetics:
A:Gene: gtfL

Query Match 44.2%; Score 3159; DB 2; Length 1449;
Best Local Similarity 47.0%; Pred. No. 8.8e-148;
Matches 669; Conservative 197; Mismatches 439; Indels 118; Gaps 28;

QY 1 MEKKYRFLKRYKRWVTVSI-----ASAVVTLTSLGSLVKADSTDDROQAVTESQAS 54
DB 1 MDKKVHYKMKVKQKQVIAVTLGSLGAVSAVSLGTNDG-VVQADEHTDATVAIPD---I 56

QY 55 LVTTSEAAKETLTATDTSTATSATSQ---PTATVTDNVSTTQSTNTTANTANFVVKPTT 111
DB 57 TVDTGTGVSNDTTAAQDPTTAAATNDVATDQATPTATPDLTDTTNTVAANAADVTATVG 116

QY 112 TSEQAKTNSDKIITTSKAVNRLTATGKFPANNTAHPKVTVD-----NNTTDTTVDRAATERRATGARRGP 166
DB 117 TDRAATT--NDTATNDTAVDTTN-----NNTTDTTVDRAATERRATGARRGP 166

QY 156 ----KIVPIKPIKIGKLOPSSLSQDDIAALGNV-----KNIRKVGKYYKYKEDGTLOKN 206
DB 167 TGGRRATPVNGTNNANNNTVTVVNDLPATNNVTDGPGSHIKTIINGKQYVVEDDGTIRKN 226

QY 207 YAL-NINGKTPFFDETGALSN-----NTLPSKGNITNNDNTNSFAQYNOVY 252
DB 227 VYLERIGGSQYFNAETGELSNGKEYRFDKNGGTGSSADSTNTNTVTVNGDKNAF-----Y 280

QY 253 STDVANFEHVHYLFAESWYRKYILKOGKTWQSTQTEKDFRPLMTWPPDQSTQYVNY 312
DB 281 GTTKDIELVDGYFTANTWYRKEILKOGKEWTASTENDKRPLLTVMWPSKAIAQASYLN 340

QY 313 MNAQ-LGHTQYNTATSLQNLAAQTIQKLEEKITAEKNTNMLRQTISAFVKTSAWN 371
DB 341 MKEQGLTNGTQTSFSSQTMQDALEVOKRIEERAREGNTDMLRTTIKNFVKTPQGN 400

QY 372 SDSEK-PFDDHLQKGLLYSNNSKLTQSANSYRILNRTPTNTQTKGKDPRYTADRTIGY 430
DB 401 STSENLDNNDHLQKGLLYNDSR-TSHANSYRLLNRTPTSTQTKHNPKYTKDTSNGGF 459

QY 431 EFLANDVNSPVVQAEQLNLHFLMFGNTYANDPDANFDSIRVDVNDVADLLQIA 490
DB 460 EFLANDIDNSPAVQAEQLNLHLYIMNIGTTTGGSEDFNFDGVRVDAVDNVAADLLQIA 519

QY 491 GDLAKAAGIHKNDKAANDHLSILEAWSYNDTPYLHDDGDNNMNDNRLSLLYSLAKP 550
DB 520 SDYFRKAKYGADQSQQA IKHLSILEAWSHNDAYNEDTKGAQLPMDDDPHLALVYLLRP 579

QY 551 LNQRSGMNPILTNSLVNRTDDNAETAAPVSPGFIKRAHDSVQDLIRNIIRTEINPNVVG 610
DB 580 IGRSGVEPLINSNDRESCKNSKRWANAYAFVRAHDSVQSIIGQIKNEINQSTGN 639

QY 611 SPTTEIKAFPIYKNDLLATEKKYTHYNALSYALLTNKSSVPRVYVYGMFTDDGQYM 670
DB 640 TPTLDEMKAFAEYKNDKMSANKQYQVNIIPSAVALMLTHKDTVPVYVYGMFTDDGQYM 699

QY 671 AHKTINYEAIETLLKARIKYVSGSGQMRNQVGNSE-----IITSRYCKGALKATD 722
DB 700 AKSPYDAIETLLKRIYRAGGQDMKVNYIGYNTNGWDAAGVLTSTVRYGTGANSASD 759

QY 723 TGDRTTRTSGVAVIEGNPSPSLRKASDRVVNNGAAHKNQAYRPLLLTNDGKIKAYHSDQ 782
DB 760 TGAETRNQGMVIVSNQPALRL--TSLNTINKGAHRNQAYRPLLLTNDGVAIYVYND 817

QY 783 EAAGLVRYTNDRGELIFTAADIKGVANPOVSGYLVGVVPGAAADQVRAASTAPSTDG 842
DB 818 DANGIKYKTDGNGNLTFANEIRGNPQVDGYLAWVPGVASENQDVRVAPSKENSSG 877

QY 843 KSVHQAALDSRVMEFGFSNFQAFATKKEEYTNVVIKAVNDKFAEWGVTDPMAAQQYVSS 902

878 LVYESNAALDSQVIYEGESNFQDFVQNPSTYNTKKIAENANLFKSWGITSFEFAQQYVSS 937
903 TDGSPFLDSVIONGYAFTDRYDLGISKPNKYGTADLVKAIKALHSKIGIKVMADWVPDQNY 962
938 DDGSPFLDSVIONGYAFTDRYDIGNSKNKGSLADLKAALKSLHVAIGISAIADWVPDQY 997

963 ALPEKEVVTATRVDKYGTVPVAGSQIKNTLYVVDVSGSKGQQAQYGGGAFLEELQAKYPL 1022
998 NLPGEVVTATRVNNYGETKDGAIIDHSLYAAKTRTFGNDYQKGGYGGAFLEELKRLYPOI 1057

1023 PARKQISTGVPMWDPVSKIKQWSAKYFNGTNIILGRGAGYVLKDQATNTYFSLVSDNTFLPK 1082
1058 FDRQVISTGKMTTDEKITQWSAKYMGNGTNIILDRGSEIVLKN-GLNGYGTNGGKVSPLK 1116

1083 SLVNPNNHGTSSSVTGLVDFDGKYY-----YYSTSGNQAKNAPISLGN-NWYTFDNN 1132
1117 -VVGSNQSTNGDNQNG--DGSCKPEKRLFSVRYRYNNGQVAKNAFIKDNDGNVYFDS 1172

1133 GYMTGAQSIINGANYFSLNGIOLRNALYDNGNKVLSVYGNDRRYENG-----1181
1173 GRMAVGEKTIIDGQYFFLANGVQLRDGYRNRGQVFFYDQNGVNLNANGKODPKPDNNNN 1232

1182 -----YVLFQGO-WRYEQ-NGIMAVGLTRVHGAVQYFDASGFOAKGOFITTAGDKLYF 1233
1233 ASGRNQFVQIGNVWAYVDGNGKRVYGHQNINGQELFPDNNNGVQVKGRTV-NENGAIYY 1291

1234 DRDSGNQISNRPVRNSKGEWFLFDHNGVAVTGTVTTFNGORLYFKPVGQVAKGEFIRDANG 1293
1292 DANGEMARNFAETEPGVMAYFNNDGTAVKGSQINGQDLYFDQNGRQVKA-LANVDG 1350

1294 YLRYYDPSNGEVRNRFVRNSKGEWFLFDHNGIATVGARVNG 1336
1351 NLRYYDPSNGELRYNR-HEIDGSIYFDGNGNAVKGVMVING 1392

RESULT 7
T30552
Glucosyltransferase N - Streptococcus salivarius (fragment)
C:Species: Streptococcus salivarius
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30552
R:Jaffe, R.I.
submitted to the EMBL Data Library, February 1998
A:Description: Streptococcus salivarius V1477 gtfN.
A:Reference number: Z20854
A:Accession: T30552
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <JAF>
A:Cross-references: UNIPROT:O68542; UNIPARC:UPI00000B10PD; EMBL:AF049609; NID:g29335545;
C:Genetics:
A:Gene: gtfN

Query Match 44.0%; Score 3146; DB 2; Length 1449;
Best Local Similarity 46.8%; Pred. No. 3.9e-147;
Matches 666; Conservative 200; Mismatches 439; Indels 118; Gaps 28;

QY 1 MEKKYRFLKRYKRWVTVSI-----ASAVVTLTSLGSLVKADSTDDROQAVTESQAS 54
DB 1 MDKKVHYKMKVKQKQVIAVTLGSLGAVSAVSLGTNDG-VVQADEHTDATVAIPD---I 56

QY 55 LVTTSEAAKETLTATDTSTATSATSQ---PTATVTDNVSTTQSTNTTANTANFVVKPTT 111
DB 57 TVDTGTGVSNDTTAAQDPTTAAATNDVATDQATPTATPDLTDTTNTVAANAADVTATVG 116

QY 112 TSEQAKTNSDKIITTSKAVNRLTATGKFPANNTAHPKVTVD-----NNTTDTTVDRAATERRATGARRGP 166
DB 117 TDRAATT--NDTATNDTAVDTTN-----NNTTDTTVDRAATERRATGARRGP 166

QY 156 ----KIVPIKPIKIGKLOPSSLSQDDIAALGNV-----KNIRKVGKYYKYKEDGTLOKN 206
DB 167 TGGRRATPVNGTNNANNNTVTVVNDLPATNNVTDGPGSHIKTIINGKQYVVEDDGTIRKN 226

Db	1154	FDGVRKAKGOSLYGSGKLTQTGTGKWEVTVKDDSGKEEFYQY-----PFKGGIMAT	1208
Qy	1200	GLTRVHGAQQVDFAGFOAKGQFIITADGKLYRFDQSGN-----	1239
Db	1209	GLTEVEGKEKYFDNGYQAKGVFVPTKDGLHMFPGDSGERKYSGFFQDGNWYANDKG	1268
Qy	1240	-----QISNRFVR-----NSKGE-----	1252
Db	1269	YVATGFTKVGKQNLVFNKGVQVGNRRFFQVGDATYANNEGDVLRAQGTINGDELYFDES	1328
Qy	1253	-----	1252
Db	1329	GKQVGEFVNNPDGTTSYDAITGVKLVDTSLVVDGQTFNVDAKGVVTKAHTPGFTYTG	1388
Qy	1253	--WFLFDHNGVAVTGTVPNGORLYFKPNQVQAKGEFIRDANGYLRYLYDPNSGNVRNR	1310
Db	1389	NNWFVADSYGRNVTAQVINGQHLVFDANGRVKGGFVTNTDGRSFFYHWTGDKLVSTF	1448
Qy	1311	VNSNGEWFPLDHNGIATVGARVNG	1336
Db	1449	FATGHRWYADDRGNVVTGAQVING	1474
RESULT 11			
A41483			
Glucosyltransferase (EC 2.4.1.-) gtfs precursor - Streptococcus sobrinus			
C:Species: Streptococcus sobrinus			
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Mar-2004			
C:Accession: A41483			
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.			
Infect. Immun. 58, 2452-2458, 1990			
A:Title: Analysis of the Streptococcus downei gtfs gene, which specifies a glucosyltransferase			
A:Reference number: A41483; MUID:90316665; PMID:2142479			
A:Accession: A41483			
A:Molecule type: DNA			
A:Residues: 1-1365 <GIL>			
A:Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:gl53652; PIDN:AAA26898.1; PID:			
C:Genetics:			
A:Gene: gtfs			
C:Keywords: glycosyltransferase; hexosyltransferase			
Query Match 40.5%; Score 2893.5; DB 2; Length 1365;			
Best Local Similarity 44.3%; Pred. No. 1e-134;			
Matches 634; Conservative 211; Mismatches 412; Indels 175; Gaps 30;			
Qy	1	MEKVRFLKRVKRWVTVSIAVAVTLTSLGSLVRADSDDDRQAAVT-----	49
Db	1	MEKNRLKLVKVKQWALGVTT--VTLFSLAGGVVAAVTNNNGTSVQVKNVPSDPK	58
Qy	50	BSQASLVTTSEAKETLTATDTSTATSATSQPTATVTDNVSTTQSTNTTANTAN-FVV	107
Db	59	FDAQAQNGLAQAMFKAANQAD-QTATSQVSPATDGRVDN-QVTPAANQPAANVANQDVA	116
Qy	108	KPTTTSQAKTNSDKLITTSKAVNRLTAT-----GKVPANNNTAHPKVTVDKIVPKP	162
Db	117	NPA-----TDGALNRQSAADTSDGKAVPQTS-----	145
Qy	163	KIGKLQPSLSQDDIAALGNVKNTRKVNKGKYYKYKEDGTGLQKNVALNINGKTFPFDETG	222
Db	146	-----QPG-----HLETVDGKTYVDANGQRLKNYSWIDGKTYYPDGTQ	185
Qy	223	ALSNNTLPKKNININNNNTNSFAQVQVYSTDVANFHVHLYLTAESWRPKYILKDGK	282
Db	186	GEAQDLP-KTGOAQDNVPDSYQANNAQVSNASSFTVDNLTADSWYRPRKILKNGQ	244
Qy	283	TWTQSTKDFRLLMTWHPDQSTQRYQVNYMAQLGIHGTNTATSPQLNLAAQTQTK	342
Db	245	SWQASSEGDLRPLMTWPPDAATKAAYANFWAKEGLISGSYRQNSA--NLDAAATQIQSA	302
Qy	343	IEEKITAENKTNWLRQTISAFVKTSQAMNSDSEK-----PFDDHLQKGLLYSNSKLT	398
Db	303	IEKKIASRGNTNWLDRKQSQFVKSQNSQMSIASENETVVPNQDHMQGGALLFS-NSKDTEH	361

RESULT 12
JC5473
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides

Qy	399	ANSNRYILNRTPTNQTGKDDPRYTADRTIGGYEFLLANDVNSNVPVQAEQLNWLFLMN	458
Db	362	ANSWRLLNRPETFTQTKG--KYFT-TNYAGYELLANDVNSNVPVQAEQLNHLHYLMN	418
Qy	459	FCNIYANDPDANFDSIRVDADVNDADLLQTAGDYLKAAGIHKNDKAANDHLSLLEAWS	518
Db	419	WGDIVMGDKDANFDCVRVDAVDNVNADLLQORDYKAKYQTDQNEKNAIDHLSLLEAWS	478
Qy	519	YNDTPYLHDDGDNMINMDNRLRLSLLSLAKPLNQRSGMNPPLITNSLVNRTDDNAETA	578
Db	479	GNNDNVYKQNNFSLSIDNDQSGMLKAFGASAYRGNLSNLATAGLKNRS-ANPDSDFV	537
Qy	579	PSYSPIRAHDSVDQLIRNIIRTEI-NPNVGY-SFTTEEIKKAEIYNKDLLATEKKYT	636
Db	538	PNYFIRAHDSVDQTRIAKIIREKLGKTNADGLTNLTDLLNKAFFDIYNQDMNATDKVY	597
Qy	637	HYNTALSALLTNKSSVPRVYVYGDMPFDDQGYMAHKTINYEAEITLLKARIKYVSGQA	696
Db	598	PNNLPMAYAWMLQNKDVTTRYVYGDYTDNGYMATKTPFYNAIETLLKGRIKYVAGQA	657
Qy	697	MRNQVGNSEIITSVRYGKGALKATDGTGRTTTSQVAVIEGNNPSLRKASDRVYVNG	756
Db	658	VSYQDWSGILTSVRYGKGANSASDAGNTETRSQSMALLINNRPNFR-AVRNLTLMNG	715
Qy	757	AAHKQAYRPLLLTNDNGIKAYHSDQEA-AGLVRYTNDRGELI FTAADIKGYANPQVSGY	815
Db	716	AAHKSQAYRPLLLSTKQGIATYLNDSVDSDRYKYTDSQGNLSFSAELQSVANAVQSGM	775
Qy	816	LGVWVPCAADQDVRVAASTAPSTDGKSVHQAALDSRVMPFEGFSNFQAPATKKEEYTN	875
Db	776	IQWVWPCAADQDVRTSPSTQATKQGNIIYHQSALDSQVIEGFSNFQAFQSDQYTN	835
Qy	876	VVIANKVDKFAEWGVTDFEMAPQVYVSSDTGSLDSVIONGYAFTDRYDLGISKPNKYGA	935
Db	836	AVIAKNGDLFKSWGITQFEMAPQVYVSSDGTFLDSVLNGYAFSDRYDLAKSKNNKYGSK	895
Qy	936	DDLKAIKALHSKGIKWADVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYYVD	995
Db	896	QDLANAIKGLQSAKIKVLSLVPNLQVNLPGKEVVVATRVNQYQQAQSGATINKTPYVAN	955
Qy	996	GKSKGQQAQYKGGAFLELOAKYFELFARKQISTGVPMDSVKIKQSAKYFNGTNIUG	1055
Db	956	TRSYG-DYQEQYGGKFLDQLYPRLFSTQISTGKFDIFSVKITNWSAKYFNGSNIUG	1014
Qy	1056	RGAGYVLKQDQATNTYPSLVSDNTFLPLKSLVNPNGHTSSVTGLVPDGGKYVYVYSTG	1115
Db	1015	RGAKYVLSE--GNKYLNLDGKFLPTVLNNTYGPQVPSANGFI SKNGGIHYLDKNGQEV	1072
Qy	1116	KNAFISLGNWYFDNNGYNTGAQSINGANYFSLNGIOLBNAIYDNGKNVLSYVNDG	1175
Db	1073	KNRFEISGSWYFSDGKMATGKTI GNDTYLFPMPNGKQLKEGYYDQKAY-YDDNG	1131
Qy	1176	RRYEN-GYLF--GO-QWRYFO-NGIMAVGLTRVHGAVQVFDASGFQAKGQFI-----	1223
Db	1132	RTWNKGFVEFRVDGQDKRWIFNGDGTIAGLVSLDNRTLYFDAYGYQVKGQTVTINGKS	1191
Qy	1224	-----	1223
Db	1192	YTFDADQGLVQTDNANPAQQAQWKLQDNGWYRKDGQLLTGTEQITDQKVPFDNG	1251
Qy	1224	-----TTADGKLYFDRSGNQISNR FVRNSKGEWFLFDHNGVATGTVTFNGQRLY	1275
Db	1252	VQVKGATDASGVLRFYDRDQHQHQYKGYSTSDNNVYVNESQVLTGLQITDQTVY	1311
Qy	1276	FKPQVQAKGFIRDANGYLRYDPSNCGNEVRNRFVRNSKGEWFLFDHNGIA	1327
Db	1312	FDDKGIQAKGKAVDENGRLRYFDADSGNMLDRW-KNVDGNWYFNRNGLA	1362

Db	703	FEMAPQVRASDSKSFDAIVQNGYAFTRDYIGNVTPKYGTAADNLLDALRALHGGQIQ	A 766
Qy	953	MADVPPQMYALPEKEVVVTAIVDKYGTPTVAGSQIKNTLYVVDGKSGKQDQAKYGGAF	L 1011
Db	763	INDWVPDQIYNLPDEQLVTAIRTDGSGDHTYGSVIDHTLYASKTVAGGIYQQ-QYGGAF	L 821
Qy	1013	EELQAKYPELPARKQISTGVPMDDPSVKIKQWSAKYFNGNTNILGRGAGYVLKQDAQNTYPS	L 1077
Db	822	EQLKATQYPLQFQQKQISTDQPMNPDIIQIKSWEAKYFNGSNIQGRGAWTVLKDQWGTQYFN	L 881
Qy	1073	LVSNTTFLPKSLVNPNGHGTSSVTVGLVDFDGGKGVYVYSGNQAKNAFISLGNNTYFNN	L 1133
Db	882	VSDAQTFELPKQLIG-----EKAKTGFVTRGKETSFYSTSGQAKSAFICDNGNWTYFDK	L 936
Qy	1133	GYMVTGAQSGINGANYFSLNSGIIQLRNA--IYDNGKNVLSYICNDGRRYENGYYLFGQQ--	L 1188
Db	937	GKMVVGQNVINGINYYFLPENGIELODAYLVHDG---MYYYNNIGKQLHNTYYQDKQFN	L 993
Qy	1189	WRYPQNGIMAVGLTRV-----HGAVQYVFDASGPOAKQFIITADGKLYRFD-----	L 123
Db	994	HYFFEDGHMAQGIVTIIQSDGTPVTQYFDENGKQKGVAVKSGDGHLYFDGASGNMLFK	L 1055
Qy	1235	-----RDSGNQISNRVFNRSKGWFLFDHNGV	L 1266
Db	1054	SWGELADGSMYVDEKGNVATGKOTINNTQVYFNDGGRQIKKNFKELADGSMYLNNGV	L 1111
Qy	1262	AVTGTVTFNGORLYFKPENGVOAKGEFIRDANGYLRYDPNSGNEVRNRFVRSKGWFLF	L 1322
Db	1114	AVTGEQIINGQTLFPGNDGRQPKTTHINATGESRYYPDSDGNMITDRFERVDGNQWAF	L 1177
Qy	1322	DHNGIATVGARVNG 1336	
Db	1174	GYDGVAVTGRIIKG 1188	
RESULT 13			
B97033			
uncharacterized protein, related to enterotoxins of other Clostridiales [import			
C;Species: Clostridium acetobutylicum			
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004			
C;Accession: B97033			
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson,			
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.			
J. Bacteriol. 183, 4823-4838, 2001			
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bac			
A;Reference number: A96500; MUID:21359325; PMID:21359325			
A;Accession: B97033			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1-2817 <KUR>			
A;Cross-references: UNIPROT:Q97K42; UNIPARC:UPI000000CA0A0; GB:AE001437; PIDN:AN			
A;Experimental source: Clostridium acetobutylicum ATCC824			
C;Genetics:			
A;Gene: CAC1079			
Query Match 6.0%; Score 428.5; DB 2; Length 2817;			
Best Local Similarity 21.0%; pred. No. 1.2e-12;			
Matches 359; Conservative 182; Mismatches 530; Indels 637; Gaps 94			
Qy	34	SLVKAOSTDRQQAUTESQASLVTTSPAKETLATDTSTATSATSOPTATVTD-NVSTT	92
Db	118	SIPVTNVTNDKNTFKNESSIN---NEA---PIPKDTSKTSSTSAQTGSDNNIPSN	170
Qy	93	NQSTNIT--ANTANFVVKPTTTTSEQA-----KTDNSDKIITTSKAV--NRLTA----	139
Db	171	NTSTNTSKNENPSNTDIKTTEAPANAFIKOTPNNQSDSALAKXKLNINLAADSSQTSK	230
Qy	140	FVPANNNTAHPK--TVTDK-----IVPIKPK-----	163
Db	231	VTSSNDA--PKVNTTSTDKKSNLNANDSQGWTKDGKYYVNGVQKGFQFSINKSIY	288
Qy	164	-----IGKLKQSPSSLSQDD-----IAALGNVK-----NIRKWNKYY	195

[illegible]

Search completed: February 11, 2006, 19:41:54
Job time : 37.3806 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:57:35 ; Search time 185.535 Seconds
(without alignments)
5228.676 Million cell updates/sec

Title: US-10-797-821-35

Perfect score: 7143

Sequence: 1 MEKKVRFPLKRVKKRWTVS.....VKVNTWILIPMKFVIM 1375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.*

1: uniprot_spot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6905	96.7	1455	1	GTFC_STRMU
2	5350	74.9	1476	1	GTFC_STRMU
3	4214.5	59.0	1590	2	Q59983_9STRE
4	4207.5	58.9	1590	2	Q55263_9STRE
5	4194	58.7	1597	1	GTFL_STRDO
6	4168.5	58.4	1592	1	GTFD_STRDO
7	3579.5	50.1	1462	1	GTFD_STRMU
8	3366.5	47.1	1577	2	Q54178_STRGN
9	3348.5	46.9	1454	2	Q69A94_LEUME
10	3293.5	46.1	1575	2	Q9LCH3_STROR
11	3216	45.0	1577	2	Q55265_STRSL
12	3197.5	44.8	1512	2	Q9WXJ5_9STRE
13	3175.5	44.5	1506	2	Q56CX8_9STRE
14	3161.5	44.3	1554	2	Q8KZL5_9STRE
15	3159	44.2	1449	2	Q55264_STRSL
16	3146	44.0	1449	2	Q68542_STRSL
17	3111	43.6	1561	2	Q58BM0_9LACO
18	3081.5	43.1	1527	2	Q8KRE1_LEUME
19	3077.5	43.1	2835	2	Q8G9Q2_LEUME
20	3070	43.0	1518	2	Q00600_STRSL
21	3069.5	43.0	1522	2	Q6TXV4_LEUME
22	3065.5	42.9	1527	2	Q9ZAR4_LEUME
23	3044	42.6	1508	2	Q52224_LEUME
24	3031	42.4	1508	2	Q9EZHS_LEUME
25	3026.5	42.4	1477	2	Q9L466_LEUME
26	2970	41.6	1330	2	Q94CN4_LEUME
27	2899	40.6	1599	2	Q00595_STRSL
28	2893.5	40.5	1365	1	GTFS_STRDO
29	2876.5	40.3	1290	2	Q48756_LEUME
30	2861	40.1	1595	2	Q58BM3_LACSK
31	2836	39.7	1338	2	Q9WXJ4_9STRE

32	2770	38.8	1463	2	Q58BM6_LACFE
33	2501.5	35.0	2057	2	Q9RE05_LEUME
34	2487	34.8	591	2	Q8VUH3_STRMU
35	2481	34.7	1016	2	Q9LCJ7_LEUME
36	2403	33.6	1772	2	Q58BN0_LACRE
37	2383.5	33.4	1772	2	Q58BN3_LACRE
38	2356.5	33.0	1781	2	Q4UCS4_LACRE
39	2351.5	32.9	1781	2	Q58BL9_LACRE
40	2351.5	32.9	1781	2	Q4JLC7_LACRE
41	1542	21.6	1619	2	Q58BM0_LACRE
42	1529	21.4	1231	2	Q58BN1_LACRE
43	1023	14.3	522	2	Q8VV10_STRSA
44	848.5	11.9	374	2	Q6ZX19_9LACO
45	428.5	6.0	2817	2	Q97K42_CLOAB

ALIGNMENTS

RESULT 1

GTFC_STRMU STANDARD: PRT: 1455 AA.
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI).
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).
GN Name=gTfC; OrderedLocusName=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f, MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P., Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 1-349.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-

CC fructose + (1,6-alpha-D-glucosyl)(n+1).

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha

CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes

CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both

CC forms of glucans.

CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.

CC -!- SIMILARITY: Contains 5 cell wall binding repeats.

CC -----

CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use as long as its content is in no way modified and this statement is not

CC removed.

CC -----

CC EMBL; M22054; AAA88592.1; -; Genomic DNA.

CC EMBL; D88652; BAA26102.1; -; Genomic DNA.

CC EMBL; D88655; BAA26106.1; -; Genomic DNA.

CC EMBL; D88658; BAA26110.1; -; Genomic DNA.

CC EMBL; D88661; BAA26114.1; -; Genomic DNA.

CC EMBL; D89978; BAA26120.1; -; Genomic DNA.

CC EMBL; AE014940; AAN58706.1; -; Genomic DNA.

CC EMBL; M17361; AAA88589.1; -; Genomic DNA.

CC PIR; J70345; J70345.

CC HSP; P06653; IH8G.

CC InterPro; IPR002479; Cell_wall_bd_put.

CC InterPro; IPR003318; Glyco_hydro_70.

CC Pfam; PF01473; CW_binding_1; 2.

CC Pfam; PF02324; Glyco_hydro_70; 1.

CC Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;

CC Transferrase.

CC SIGNAL

FT CHAIN 1 34

FT REPEAT 1126 1159 Glucosyltransferase-SI.

FT REPEAT 1126 1159 A repeat.

FT REPEAT 1169 1200 A repeat.

FT REPEAT 1227 1238 C repeat.

FT REPEAT 1253 1303 AC repeat.

FT REPEAT 1318 1330 A repeat (incomplete).

FT REGION 35 1050 Catalytic (approximate).

FT REGION 1126 1455 2,4 A, 1 C and 1 AC repeats.

FT REGION 1126 1455 Glucan-binding (approximate).

FT VARIANT 21 21 V -> I (in strain GS-5).

FT VARIANT 81 81 P -> L (in strain MT4239).

FT VARIANT 106 106 D -> V (in strain GS-5).

FT VARIANT 116 116 S -> A (in strain GS-5 and strain MT4467).

FT VARIANT 126 126 A -> T (in strain GS-5).

FT VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239 and strain MT4467).

FT VARIANT 256 256 A -> V (in strain GS-5 and strain MT4467).

FT VARIANT 425 425 R -> N (in strain MT4251).

FT VARIANT 519 519 Y -> D (in strain MT4245 and strain MT4251).

FT VARIANT 538 538 R -> K (in strain MT4245 and strain MT4251).

FT VARIANT 545 545 Y -> F (in strain MT4245 and strain MT4251).

FT VARIANT 597 597 N -> D (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT VARIANT 600 600 R -> K (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT VARIANT 601 601 A -> T (in strain GS-5).

FT VARIANT 727 727 M -> T (in strain MT8148).

FT VARIANT 734 734 T -> I (in strain MT8148).

FT VARIANT 964 964 A -> V (in strain MT8148).

FT VARIANT 1113 1113 L -> F (in strain MT4239).

FT VARIANT 1118 1118 N -> Y (in strain MT4239).

FT VARIANT 1204 1204 I -> V (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).

FT VARIANT 1208 1208 V -> I (in strain MT8148).

FT VARIANT 1292 1294 DGH -> NGY (in strain GS-5, strain MT4467

FT VARIANT 1305 1369 and strain MT8148).

FT VARIANT 1326 1326 Missing (in strain MT4245).

FT VARIANT 1331 1331 I -> V (in strain MT8148).

FT VARIANT 1377 1377 T -> A (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).

FT VARIANT 1398 1398 R -> K (in strain MT8148).

FT VARIANT 1424 1424 V -> I (in strain MT8148).

FT VARIANT 1439 1439 D -> N (in strain MT4239).

FT VARIANT 1444 1444 V -> I (in strain MT4239 and strain MT8148).

FT VARIANT 1455 1455 S -> P (in strain MT8148).

FT CONFLICT 1337 1455 ORLYFKNGVQAKGRIKIVYDPNSGNEVNRNVVR

FT TSSGNGWYFPGDGYALIGHVVEGRVDFDENGVIYASHD

FT QRNHWDYDRDQFGRSSAVRFSRNGDFDNFRFF ->

FT HASILSLMVFRLRSLSSQSVKVSNTMILIPMKFVIM

FT (in Ref. 1).

SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A99A4FEC86 CRC64;

Query Match 96.7%; Score 6905; DB 1; Length 1455;

Best Local Similarity 99.0%; Pred. No. 0;

Matches 1323; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 MEKKVRFKLRKVKRWTVSIASAVVLTLSLGSGLVKADSTDDROQAVTESQASLVTTSE 60

DB 1 MEKKVRFKLRKVKRWTVSVASAVVLTLSLGSGLVKADSTDDROQAVTESQASLVTTSE 60

QY 61 AAKETLTATDTSTATSATSQPTATVTDNVSTTQSTNTTANTANFVVKPTTTTSEQAKTDN 120

DB 61 AAKETLTATDTSTATSATSQPTATVTDNVSTTQSTNTTANTANFVVKPTTTTSEQAKTDN 120

QY 121 SDKIITTSKANRLTATGKFPANNNTAHPKTVTDKIPIKPKIGKLGKQPSLSQDDIAA 180

DB 121 SDKIITTSKANRLTATGKFPANNNTAHSRTVDKIPIKPKIGKLGKQPSLSQDDIAA 180

QY 181 LGNVKNIRKVGKYYKEDGTLQKNYALNTNGKTFEDETALSNNTLPSKKGNTTND 240

DB 181 LGNVKNIRKVGKYYKEDGTLQKNYALNTNGKTFEDETALSNNTLPSKKGNTTND 240

QY 241 NTNSPAQYNQVYSTDVANFEVDHYLTAEVWPYKYLKDGKWTQTSTKQKDFRPLMTWM 300

DB 241 NTNSPAQYNQVYSTDAAFEVDHYLTAEVWPYKYLKDGKWTQTSTKQKDFRPLMTWM 300

QY 301 PDQETQROYVNMNAQLGIHOTYNTATSPLOLNLAQIOTKIEBKITAETKNTWLRTI 360

DB 301 PDQETQROYVNMNAQLGIHOTYNTATSPLOLNLAQIOTKIEBKITAETKNTWLRTI 360

QY 361 SAFVKTSQAWNSDSEKPPDDHLQKALYSNNKLTQANSNYRILNRTPTNQTGKQDPR 420

DB 361 SAFVKTSQAWNSDSEKPPDDHLQKALYSNNKLTQANSNYRILNRTPTNQTGKQDPR 420

QY 421 YTADRTIGGYEFLANDVNSNPVVOAQEQLNWLHFLMNFNFIYANDPDANFDSIRDAVD 480

DB 421 YTADRTIGGYEFLANDVNSNPVVOAQEQLNWLHFLMNFNFIYANDPDANFDSIRDAVD 480

QY 481 NVADALLOIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDTPYLHDDGDNMNDNRLR 540

DB 481 NVADALLOIAGDYLKAAGIKHNDKAANDHLSILEAWSYNDTPYLHDDGDNMNDNRLR 540

QY 541 LSLLYSLAKPLNQRSGMPLTINSLVNRTDQNAETAAVPSYSFIRAHDSQVODLIRNIIR 600

DB 541 LSLLYSLAKPLNQRSGMPLTINSLVNRTDQNAETAAVPSYSFIRAHDSQVODLIRNIIR 600

QY 601 TEINENVVGYSTFTBEEIKKAFIEYNKOLLATEKKYTHYNTALSYALLLTNKSQSVPRVYVG 660

DB 601 AEINENVVGYSTFTBEEIKKAFIEYNKOLLATEKKYTHYNTALSYALLLTNKSQSVPRVYVG 660

QY 661 DMFTDDGGYMAHKTINYEAITLLKARIKYVSGGAMRNQOVGNSEIITSVRYGKALKA 720

DB 661 DMFTDDGGYMAHKTINYEAITLLKARIKYVSGGAMRNQOVGNSEIITSVRYGKALKA 720

QY 721 TDTGDRTRTSGVAVIEGNNPSLRKASDRVVVNNGAHKNQAYRPLLLTTDNGIKAYHS 780

DB 721 TDTGDRTRTSGVAVIEGNNPSLRKASDRVVVNNGAHKNQAYRPLLLTTDNGIKAYHS 780

QY 781 DQAAAGLVRYNDRGELIFTAADIKGYANPQVSGYLGWVVPVGAADQDVRVAAPST 840
 DB 781 DQAAAGLVRYNDRGELIFTAADIKGYANPQVSGYLGWVVPVGAADQDVRVAAPST 840
 QY 841 DGKSVHONALDSRVNFECSFNFQATKKEEYTNVIAKNDKFAEWGTFDEMAPOYV 900
 DB 841 DGKSVHONALDSRVNFECSFNFQATKKEEYTNVIAKNDKFAEWGTFDEMAPOYV 900
 QY 901 SSTGDSFSLDSVIONGVAFTDRYDLGISKPNKYGTADDLVKAIKALSKGIKWADWVPDQ 960
 DB 901 SSTGDSFSLDSVIONGVAFTDRYDLGISKPNKYGTADDLVKAIKALSKGIKWADWVPDQ 960
 QY 961 MYALPEKEVVTATRVNDKYGTPVAGSQIKNTLYVVDGKSGKQQAQYGGAFLEELQAKYP 1020
 DB 961 MYALPEKEVVTATRVNDKYGTPVAGSQIKNTLYVVDGKSGKQQAQYGGAFLEELQAKYP 1020
 QY 1021 ELFPARKQISTGVPMDSVKIKOWSAKYFNGTINILGRGAGYVLKQDATNTYFSLVSDNTEFL 1080
 DB 1021 ELFPARKQISTGVPMDSVKIKOWSAKYFNGTINILGRGAGYVLKQDATNTYFSLVSDNTEFL 1080
 QY 1081 PKSLVNPNGHTSSVGLVDFDGKGYVYVYSTSGNOAKNAFISLGNWYFDNNGYVMTGAQ 1140
 DB 1081 PKSLVNPNGHTSSVGLVDFDGKGYVYVYSTSGNOAKNAFISLGNWYFDNNGYVMTGAQ 1140
 QY 1141 SINGANYFSLNGIQLRNAIYDNGNKVLSYGGNDGRRYENGYYLFGQWRIFYQNGIMAYG 1200
 DB 1141 SINGANYFSLNGIQLRNAIYDNGNKVLSYGGNDGRRYENGYYLFGQWRIFYQNGIMAYG 1200
 QY 1201 LTRVHGAOVYFDASGFOAQKQFITTADGKLRIFDRDSGNQISNRFVNRNSKGWFLFDHNG 1260
 DB 1201 LTRVHGAOVYFDASGFOAQKQFITTADGKLRIFDRDSGNQISNRFVNRNSKGWFLFDHNG 1260
 QY 1261 VAVTGTVTENGRLFKPKNGVQAKGEFIRDANGLYRLYDPNSGNEVRNRFVNRNSKGWFL 1320
 DB 1261 VAVTGTVTENGRLFKPKNGVQAKGEFIRDANGLYRLYDPNSGNEVRNRFVNRNSKGWFL 1320
 QY 1321 FDHNGIATVGARVAVNG 1336
 DB 1321 FDHNGIATVGTRVAVNG 1336

RESULT 2

GTFFB_STRMU
 ID GTFFB_STRMU STANDARD; PRT: 1476 AA.
 AC P08987; O69381; O69384; O69387; O69390; O69396;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 GN Name=gtfB; OrderedLocusName=SMU.1004;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RP [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=87308013; PubMed=3040685;
 RA Shiroza T., Ueda S., Kuramitsu H.K.;
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
 RL J. Bacteriol. 169:4263-4270(1987).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
 RX MT4467 / Serotype e, and MT8148 / Serotype c;
 RX MEDLINE=98231643; PubMed=9570124;
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
 RA Kimura S., Hamada S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEWS Microbiol. Lett. 161:331-336(1998).

RN [3]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carlson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
 RT pathogen."; Acad. Sci. U.S.A. 99:14434-14439(2002).
 RL Proc. Natl.
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to
 CC play a key role in the development of the dental plaque because of
 CC their ability to adhere to smooth surfaces and mediate the
 CC aggregation of bacterial cells and food debris.
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
 CC fructose + (1,6-alpha-D-glucosyl)(n+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
 CC forms of glucans.
 CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
 CC -!- SIMILARITY: Contains 10 cell wall binding repeats.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC
 CC EMBL, M17361; AAA88588.1; -; Genomic DNA.
 CC EMBL, D88651; BAA26101.1; -; Genomic DNA.
 CC EMBL, D88654; BAA26105.1; -; Genomic DNA.
 CC EMBL, D88657; BAA26109.1; -; Genomic DNA.
 CC EMBL, D88660; BAA26113.1; -; Genomic DNA.
 CC EMBL, D89977; BAA26119.1; -; Genomic DNA.
 CC EMBL, AB014940; AAN58705.1; -; Genomic DNA.
 CC PIR, B33135; B33135.
 CC HSP, P06653; IH8G.
 CC InterPro; IPR002479; Cell_wall_bd_put.
 CC InterPro; IPR003318; Glyco_hydro_70.
 CC Pfam; PF01473; CW_binding_1; 4.
 CC Pfam; PF02324; Glyco_hydro_70; 1.
 KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
 KW Transferrase.
 FT SIGNAL 1 34 Potential.
 FT CHAIN 35 1476 Glucosyltransferase-I.
 FT REPEAT 1097 1130 A repeat.
 FT REPEAT 1161 1210 1.
 FT REPEAT 1225 1275 2.
 FT REPEAT 1290 1340 3.
 FT REPEAT 1355 1405 4.
 FT REPEAT 1420 1470 5.
 FT REGION 35 1051 Catalytic (approximate).
 FT REGION 1097 1476 Glucan-binding (approximate).
 FT REGION 1161 1470 5 X tandem repeats.
 FT VARIANT 62 62 S -> T (in strain MT4239).
 FT VARIANT 65 65 T -> I (in strain GS-5).
 FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245,
 FT strain MT4251, strain MT4467 and strain
 FT MT8148).
 FT VARIANT 78 78 Q -> P (in strain MT4251).
 FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245,
 FT strain MT4251, strain MT4467 and strain
 FT MT8148).
 FT VARIANT 89 89 S -> F (in strain MT4251).
 FT VARIANT 168 168 K -> N (in strain MT4251).
 FT VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and
 FT strain MT8148).
 FT VARIANT 399 399 N -> R (in strain MT4239).
 FT VARIANT 474 474 I -> T (in strain MT4239).
 FT VARIANT 512 512 K -> R (in strain MT8148).

FT	VARIANT	519	519	F -> Y (in strain MT8148).
FT	VARIANT	701	701	T -> I (in strain MT8148).
FT	VARIANT	708	708	A -> V (in strain MT8148).
FT	VARIANT	938	938	F -> L (in strain MT8148).
FT	VARIANT	952	957	FGKPYE -> YGTPVA (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	963	964	SV -> NT (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	968	970	ADS -> VDG (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	1086	1086	A -> T (in strain MT4239).
FT	VARIANT	1158	1158	S -> N (in strain MT4239).
FT	VARIANT	1163	1163	H -> Y (in strain MT4251).
FT	VARIANT	1168	1168	E -> K (in strain MT8148).
FT	VARIANT	1182	1182	Y -> C (in strain MT8148).
FT	VARIANT	1234	1234	A -> P (in strain MT4239).
FT	VARIANT	1263	1263	R -> H (in strain GS-5 and strain MT4467).
FT	VARIANT	1263	1263	R -> P (in strain MT8148).
FT	VARIANT	1264	1264	Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1272	1272	S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1329	1329	H -> Y (in strain GS-5 and strain MT4467).
FT	VARIANT	1394	1394	Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1402	1402	S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1459	1459	Y -> H (in strain MT4467).
FT	CONFLICT	570	570	R -> A (in Ref. 1).
FT	CONFLICT	800	817	AQDVRVAASTAPSTDGK -> LIXMFALRLARPHQWA (in Ref. 1).
FT	CONFLICT	1310	1310	H -> L (in Ref. 1).
FT	CONFLICT	1476	1476	MM; 9C6E09F731B4BCFC CRC64;
SQ	SEQUENCE	1476	1476	AA; 165847 MM; 9C6E09F731B4BCFC CRC64;
Query Match				
Best Local Similarity 74.9%; Score 5350; DB 1; Length 1476;				
Matches 1038; Conservative 94; Mismatches 168; Indels 44; Gaps 11;				
QY	1	MEKVKRFLKRVKQKVVTVSIA	SVVTLTSLGSLVKADSTDDRQAVTESQASLVTTSE	60
DB	1	MDKVKRYKRVKQKVVTVSIV	ASAVMTLTLSGLVKADSNESKQISNDSTNTVTANE	60
QY	61	AAKETLTATDTATATSQPTAT	VDNVSTNQSTNTANTANFVVKPTTTSBAQKTDN	120
DB	61	--ESNVTTVTSKQEAASQTNH	T-----TTISSSTSVNPKVSNPYTVGETA--SN	111
QY	121	SKIIITTSKANRLTATGKFPAN	NNTAHPKVTVDKIVIPKIGKIKQPSLSODDIAA	180
DB	112	GEKLNQITTVDTKSEA-----	AANNISKQTTED-----TDVIDDSNAA	151
QY	181	----LGNVKNIRKVGKYYYKED	GTLOKNYALNINNGKTFPPDETGALSNNLTSPKKGNI	236
DB	152	NIQILEKLPNVKEIDGKYYYD	NNGKVRTNFTLIADGKILHFDGTAYDTSIDTVNKDI	211
QY	237	TNNDNNSFAQNVYSTVDVAN	FEVHDHYLTAESWRPKYILKDGKTWTQSTEKDFRPLL	296
DB	212	VTT-RSNLYKYNQVYDRSAQ	FEVHDHYLTAESWRPKYILKDGKTWTQSTEKDFRPLL	270
QY	297	MTWMPDQETOROVYNNMNA	QLGIHOTYNTATSPLOLNLAQTIOTKIEKTAENKNWL	356
DB	271	MTWSPSETQRYVYNNMNA	QLGINKTYDDTSNQLQNLNIAAATIOAKIEAKITTLKNNDWL	330
QY	357	RTQISAFVKTSAMNSDSKPD	DDHLQKGALLYSNNSKLTSQANSNYRILNRPNTQTKG	416
DB	331	RTQISAFVKTSAMNSDSKPD	DDHLQNGAVLYDNEGKLTTPYANSNYRILNRPNTQTKG	390
QY	417	KDPRYTADRTIGGYBFLAN	VDNSNPVVQABQLNWLHFLMNFIGNIYANDPDANFDSIRV	476
DB	391	KDPRYTADRTIGGYBFLAN	VDNSNPVVQABQLNWLHFLMNFIGNIYANDPDANFDSIRV	450
QY	477	DAVDNVDADLLQIAGDYL	KAAGIHKNDKKAANDHLSILEAWSYNDTPYLLHDDGDNMINWD	536

DB	451	DAVDNVDADLLQIAGDYL	KAAGIHKNDKKAANDHLSILEAWSYNDTPYLLHDDGDNMINWD	510
QY	537	NRLRLSLYSIAKPLNQBSGN	NPLITNSLVNRTDDNAETAAPVPSYSFIRAHDSYVODLIR	596
DB	511	NKRLSLFLSLAKPLNQBSGN	NPLITNSLVNRTDDNAETAAPVPSYSFIRAHDSYVODLIR	570
QY	597	NIIRTEINPNVGVSTTEEB	EIKKAFIYNKOLLATEKKYTHYNTALSALLITNKSSVPR	656
DB	571	DIIRKABINPNVGVSTFTEE	EIKKAFIYNKOLLATEKKYTHYNTALSALLITNKSSVPR	630
QY	657	VYVGDMFTDDGOYMAHKT	INTYEAETLLKARIKYVSGQAMRNQOVGNSEIITSVRYGKG	716
DB	631	VYVGDMFTDDGOYMAHKT	INTYEAETLLKARIKYVSGQAMRNQOVGNSEIITSVRYGKG	690
QY	717	ALKATDTGDRTRTSGVA	VIENGNPSRLKASDRVVNNGAAHKQAYRPLLLTTDNGIK	776
DB	691	ALKATDTGDRTRTSGVA	VIENGNPSRLKASDRVVNNGAAHKQAYRPLLLTTDNGIK	750
QY	777	AYHSDQEAAGLVRYTND	RGELIFTAAADIKGYANPOVSGYLGWVPVGAADODVRAAST	836
DB	751	AYHSDQEAAGLVRYTND	RGELIFTAAADIKGYANPOVSGYLGWVPVGAADODVRAAST	810
QY	837	APSTDGKSVHQNAAAL	DSRVMPFEGFSNFQAFATKKEEYTNVVIKNDVKFAEWGVTDFEMA	896
DB	811	APSTDGKSVHQNAAAL	DSRVMPFEGFSNFQAFATKKEEYTNVVIKNDVKFAEWGVTDFEMA	870
QY	897	PQYVSTSDGSLDSVION	GYAFTDRYDILGISKPNKYGTADDLVKAIKALHSKGIKVMADW	956
DB	871	PQYVSTSDGSLDSVION	GYAFTDRYDILGISKPNKYGTADDLVKAIKALHSKGIKVMADW	930
QY	957	VPDQMYALPEKEVVTAT	RVDKYGTPVAGSOIKVTLVYVDGKSSGKQQAQKYGGAFLLEIQ	1016
DB	931	VPDQMYALPEKEVVTAT	RVDKYGTPVAGSOIKVTLVYVDGKSSGKQQAQKYGGAFLLEIQ	990
QY	1017	AKYPELPARKOISTGV	PMDPSVKIKOWSAKVFNGTNILGRGAGYVLKQOATNTYFSLVSD	1076
DB	991	AKYPELPARKOISTGV	PMDPSVKIKOWSAKVFNGTNILGRGAGYVLKQOATNTYFSLVSD	1049
QY	1077	N---TFLPKSLVNP	NHGTSSSVTGLVFDGKGYVYVYSTSGNQAKNAFISLGNMNYFDNNG	1133
DB	1050	NKEINFLPKTLN-----	QDSQVGSYDGKGYVYVYSTSGYQAKNTFISEGDKWYFDNNG	1104
QY	1134	YMTVTAQOSINGANY	YFLSNGIQLRNAIYDNGKVLVSYGNDGRRYENGYYLF-CQOWRYF	1192
DB	1105	YMTVTAQOSINGANY	YFLSNGIQLRNAIYDNGKVLVSYGNDGRRYENGYYLF-CQOWRYF	1164
QY	1193	ONGIMAVGLTRVHGAV	QYFDASGFOAKGQFITTADGKLYRFDSDSGNOISNRFVNSKGE	1252
DB	1165	NGEMSUGLTVIDGO	VYQYFDEMGYQAKGFVTTADGKLYRFDKQSGNMYRNFRIENEGK	1224
QY	1253	WFLFDHNGAVTGT	VTFNGQRLYFKPNQVQAKGEFIRDANGYLYRYDDPNSGNEVNRNFRV	1312
DB	1225	WLYLGEDAAVTGS	QTINGQHLVFRANGVQKGEFVTDYGRISYYDSNSGDIQRNFRV	1284
QY	1313	NSKGEWFLFDHNG	TAVTGARVUNG 1336	
DB	1285	NAQQWFFYDNN	GYAVTGARTING 1308	
RESULT 3				
Q59983 9STRE PRELIMINARY; PRT; 1590 AA.				
ID	Q59983 9STRE			
AC	Q59983;			
DT	01-NOV-1996 (TReMBLrel. 01, Created)			
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)			
DT	01-MAR-2004 (TReMBLrel. 26, Last annotation update)			
DE	Glucosyltransferase-I precursor (EC 2.4.1.5).			
GN	Name=gcf1;			
OS	Streptococcus sobrinus.			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1310;			

produced from *Streptococcus sobrinus* ATCC 33478.²;
Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).

[2]

PROTEIN SEQUENCE.
MEDLINE=91224988; PubMed=1827439;
Mooser G., Heita S.A., Paxton R.J., Shively J.E., Lee T.D.;
"Isolation and sequence of an active-site peptide containing a
catalytic aspartic acid from two *Streptococcus sobrinus* alpha-
glucosyltransferases.";
J. Biol. Chem. 266:8916-8922(1991).
EMBL; D63570; BAA09792.1; -; Genomic_DNA.

PIR; A39841; A39841.

HSP; P06653; LGVM.

GO; GO:0009250; P-glucan biosynthesis; IEA.

InterPro; IPR002479; CW_binding.

InterPro; IPR003318; Glyco_hydro_70.

Pfam; PF01473; CW_binding_1; 3.

Pfam; PF02324; Glyco_hydro_70; 1.

SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;

Query Match 58.9%; Score 4207.5; DB 2; Length 1590;

Best Local Similarity 60.2%; Pred. No. 3.4e-192;

Matches 810; Conservative 185; Mismatches 298; Indels 53; Gaps 14;

```

QY 1 MEKVRFKLRKVRKRWTVSIAVAVTLTSLGSLVKAD---STDRQQAUTESQASLVT 57
DB 1 MEKVRFKLRKVRKRWTVSIAVAVTLTSLGSLVKAD---STDRQQAUTESQASLVT 58
QY 58 TSEAAKETLTATDTSTATSQPTATVDNVTSTNQSTNTTANTANFVVKPTTSEQAK 117
DB 59 NQOATDQTSIA---ATATSEQASDAATDQASAAEQTQGTAST-DTAAQTITNANEA- 113
QY 118 TNSDKIITTSKAVNRLTATGKFPVANNTHPKVTVDKIVPKIGIKLKOPSSLSQDD 177
DB 114 -----KWPPTENEN---QGPTDEM-----LAEAKNVATAESDS 143
QY 178 IAA-LGNVKNIRKNGKYYYKXEDGTLQKNYALNNGTKTFPDETCALSNNT-LPSKGN 235
DB 144 FPSDLAAGSNVQVQDKYIYYDDGNVKNKFAVSQDKIYIFDETGAAYKQDTSKVADKSS 203
QY 236 ITNNNTNSFAQYNOVYSTDVANFEHVDHYLTAEWSYRPKYILKDGKTWTQSTKDFRPL 295
DB 204 SAVSQNATIFAANNRAYSTSAENFEAVDNYLTADSWYRPKSILKDGKTWTESKDFRPL 263
QY 296 LMTWPPDQRTQRYNNYMAQLGIGHTYNTATSPILNLAAQTIQTKIBEKITAENQTNW 355
DB 264 LMAWPPDTEKRYNNYNNLVVIGIDKITYAETSAQDLTAAAEILVQARIEQKITTEQNTKW 323
QY 356 LRQTISAFVKTSANSDSEKPPDDHLOKGLLYSNNSKLTQOANSYRILARTPTNQTG 415
DB 324 LREAISAFVKTPQWNGSEKPYDDHLOKGLKLFQNSDLTPDTQSNYRLLARTPTNQTG 383
QY 416 KKDPRYT--ADRTIGYBELLANDVNSNPVQAEQLNWLHFLMFGNIYANDPDANFDS 473
DB 384 SLDSRTYNANDPLGGYFELLANDVNSNPVQAEQLNWLHFLMFGSIYAKDADANFDS 443
QY 474 IRVDAVDNVADLLQIAGYLLKAAKGIHKNKDAANDHLSILBAWSYNDTPYLLHDDGNMI 533
DB 444 IRVDAVDNVADLLQISDYLKAAVGIDKNNKANNHVSIVEAWSNDTPYLLHDDGNILM 503
QY 534 NNDNRLRLSLYLAKPLNQRSGMPLITNSLVNRTDQNAETAAPVSYSFIRAHSEVOD 593
DB 504 NMDNFRSLWSLAKPLDKRSLGNPLIHNSLVDRVDDRETVETVPSYSFAAHSEVOD 563
QY 594 LIIRIRTEINPNVNGYSPTTEIKKAFIYNKDLATEKKYTHVNTALSVALLLTNKSS 653
DB 564 IIRDIKAEINFSFGYSFTQEBIDQAFKIYNEDLKKTKYTHVNPVLSYLLTNKGS 623
QY 654 VPRVYGDFTDDGGYMAHKTNYEATILLKARIKYVSGGQARNQOVNGSEILTSVRY 713
DB 624 IPRVYGDFTDDGGYMAHKTNYEATILLKARIKYVSGGQARNQOVNGSEILTSVRY 683
QY 714 GKGLKATDGTDRTRTSGVAVIEGNNPSLRKASDRVVVNVGAAHKNQAYRPLLLTTDN 773

```

```

DB 684 GKGLKQSDKGDATRTTSVGVMGNQNFSLDGK-VVALNNGAAHQAQYERALAVSTKD 742
QY 774 GIKAYHSDQEA--AGLVRYTNDRGELIFTAADIKIGYANPOVSGYLGWVPVPGAAADQDVR 831
DB 743 GVATYATDADASKAGLVRTDENGLYFLNDDLKGVANPOVSGFLQVWVPVPGAADQDTR 802
QY 832 VAASAPSTDGKSVHONAAALDSRVMEFGFSNFOAEATKKEEYTNVVIKAVNDKFAEWGVT 891
DB 803 VAASDTASTDGKSLHQDAMDSRVMEFGFSNFOAEATKKEEYTNVVIKAVNDKFAEWGVT 862
QY 892 DPEMAPOVSVSDGSLDSVITQNGYAFTRDYLGIKSPKNKYGTADDLKAIKALHAKGKIK 951
DB 863 DPEMAPOVSVSDGSLDSVITQNGYAFTRDYLGIKSPKNKYGTADDLKAIKALHAKGKIK 922
QY 952 VMADWVPDQMYALPEKVVATRVKGYTPVAGSQIKNTLYVVDGKSGKQDQAKYGGAF 1011
DB 923 VMADWVPDQMYALPEKVVATRVKGYTPVAGSQIKNTLYVVDGKSGKQDQAKYGGAF 982
QY 1012 LEELOQAKYPELFARKQIISTGVPMDSVKIKOWSAKYFNGTILGRGAGYVLKDAQATNTYF 1071
DB 983 LDELKEKYPELFARKQIISTGVPMDSVKIKOWSAKYFNGTILGRGAGYVLKDAQATNTYF 1042
QY 1072 SILVSDNTLPSLVNPNHGTSSVTGLVFDGKGYVYVYST-SGNQAKNAFISLGNWYTFD 1130
DB 1043 NVASDTLFLPSLLG-----KWVESGIRYDGKGYLYNSSATGDQVKAFITEAGNLYYFG 1097
QY 1131 NNGYVMTGAQINGANYFLSNGIQLRNAIYDNGHKNVLSYSGNDGRRYENGYYLFGQOQR 1190
DB 1098 KDGWVMTGAQTINGANYFLSNGIQLRNAIYDNGHKNVLSYSGNDGRRYENGYYLFGQOQR 1157
QY 1191 YFQNGIMAVGLTRVHGAVOYFDPASFOAKGQFIITADGKLYRFDSDSGNOISNRFVRNSK 1250
DB 1158 YFQNGIMAVGLTRVHGAVOYFDPASFOAKGQFIITADGKLYRFDSDSGNOISNRFVRNSK 1217
QY 1251 GEWFLFDHNGVAVTGTVTFTNGORLYFKPENGVOAKGEFIRDANGYLYRYPDPSNGEVRNRF 1310
DB 1218 GHWYLYLKGQVAVTGAQTVGKQKLYFEANGEQVKDFTVSHGKLYFYDVSDDMWTDTF 1277
QY 1311 VRNSKGEWFLFDHNGIATGARVNG 1336
DB 1278 IEDKAGNWFYLGKDGAAVSGAQTIRG 1303

```

RESULT 5

GTFL STRDO

ID GTFL STRDO STANDARD; PRT; 1597 AA.

AC P11001;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 13-SEP-2005 (Rel. 48, Last annotation update)

DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)

DE (Sucrose 6-glucosyltransferase).

GN Name=gtfI;

OS *Streptococcus downei* (*Streptococcus sobrinus*).

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC *Streptococcus*.

ON NCBI_TaxID=1317;

RX [1]

RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].

RC STRAIN=MF28;

RX MEDLINE=87308014; PubMed=3040686;

RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;

RT "Nucleotide sequence of a glucosyltransferase gene from *Streptococcus*

sobrinus MF28.";

RL J. Bacteriol. 169:4271-4278(1987).

CC -!- FUNCTION: Production of extracellular glucans, that are thought to

play a key role in the development of the dental plaque because of

their ability to adhere to smooth surfaces and mediate the

aggregation of bacterial cells and food debris.

CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-

fructose + (1,6-alpha-D-glucosyl) (n+1).

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -|- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
 CC forms of glucans.
 CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -|- SIMILARITY: Contains 19 cell wall binding repeats.
 CC -----
 CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL: M17391; AAC63063.1; -; Genomic DNA.
 DR InterPro: IPR002479; Cell wall bd put.
 DR InterPro: IPR003318; Glyco_hydro_70.
 DR Pfam: PF01473; CW binding 1; 4_
 DR Pfam: PF02324; Glyco_hydro_70; 1.
 KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
 FT SIGNAL 1 38 Potential.
 FT CHAIN 39 1597 Glucosyltransferase-I.
 FT REPEAT 1099 1132 A repeat.
 FT REPEAT 1163 1213 AC repeat.
 FT REPEAT 1229 1277 AC repeat.
 FT REPEAT 1292 1342 AC repeat.
 FT REPEAT 1352 1399 B repeat.
 FT REPEAT 1406 1455 AC repeat.
 FT REPEAT 1465 1512 B repeat.
 FT REPEAT 1519 1568 AC repeat.
 FT REPEAT 1582 1597 A repeat (incomplete).
 FT REGION 39 1050 Catalytic (approximate).
 FT REGION 1099 1597 1.25 A, 2 B and 5 AC repeats.
 FT REGION 1099 1597 Glucan-binding (approximate).
 SQ SEQUENCE 1597 AA; B9E86A200868798E CRC64;

Query Match 58.7%; Score 4194; DB 1; Length 1597;
 Best Local Similarity 59.6%; Pred. No. 1.5e-191;
 Matches 807; Conservative 186; Mismatches 299; Indels 62; Gaps 14;
 QY 1 MEKVRFLKVRKVRWTVSYIASAVVLTSLSGSLVKAD-----STDROQAVTESQASLV 56
 DB 1 MEKNERFQHKVRKVRWTVSYIASATMLASALGASVASADTEVTSEDSNQAVTADQTTN 60
 QY 57 TTSEAKETLTATDTSTATSQTATVTDNVSTTNTQNTANTANFVPTTTTSEQA 116
 DB 61 ODTEQTSVAATATSEQASTDAATQASATDQASAAEQTQGTAST-DTAAQTNTANEA 119
 QY 117 K---TDNSDKIT---TSKAVRLTATGKFPVANNNTAHPKVTVDKIVIPKIKGLKOP 170
 DB 120 KVVPTENENQVTDMLAEAKNVTAESNSIPS----- 152
 QY 171 SSLSQDDTAALGNKVRKVVYKEDGTLQKNYALNKGTFPPDETGALSNNTL- 229
 DB 153 -----DLAAMSNV----QVDGKYDDGNNKVFVSGEKIYFDETGAIKDTSKV 203
 QY 230 -PSKGNITNNDNTNSFAQYNQVSYTDVANFEHVDHLYTAESWYRKYIKDKGKTWOST 288
 DB 204 EADKSGSDISKET-TFAANNRAYSTSAENFEADNYLTADSWYRPSKILDKGKTWESS 262
 QY 289 EKDFRPLMTWPDQETQRYNNYNAQLGIIHQNTNTATPLQLNLAQTIOKIEKIT 348
 DB 263 KDDFRPLMAWPDTEKRYNNYNNKVVGIDKTYTAETSQADLTAAAEVLQARIEKIT 322
 QY 349 AEKNTNWLRLQTSASFVKTSANNSSEKPPDHLQKALLYSNNKLTQSANNRYLNR 408
 DB 323 TEQNTKWLREALSAFVKTPQWNGSEKPYDDHLQNGALKFNQSDLTPTDTSNRYLNR 382
 QY 409 TPTNQTGKDPRT--ADRTIGCYEFLANDVNSNPVQASQLNWLHFLMNFNIYAND 466
 DB 393 TPTNQTGSLDSFTYANDPLGGYELLALLANDVNSNPVQASQLNWLHFLMNFNIYAKD 442
 QY 467 PDANFDSIRVDVNDVADLLQIAGDYLKAAKGIIHKNDKKAANDHLSILEAMSYNDTPYLH 526
 |||||

DB 443 ADANFDSIRVDVNDVADLLQIASSDYLKAAAGIDKNNKNNANNHVSIVEAWSNDTPYLH 502
 QY |||||
 DB 527 DDCDNMNNDRRLRLSLLYSLAKPLNQRSGMPLITNSLVNRDNDNAETAAPSYSFIRA 586
 DB 503 DGDGDLNMNDKFRSLWSLAKPLDKRSLAPLIHNSLVDRVDREVETVPSYFARA 562
 QY HDSVQDLIRDIIRIIRTEINPNVVGYSFTTEIKKAFIYNKDLATEKKYTHYNALSYAL 646
 DB 563 HDSVQDLIRDIIRIIRTEINPNVVGYSFTTEIKKAFIYNKDLATEKKYTHYNALSYAL 622
 QY 647 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 706
 DB 623 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 682
 QY 707 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 766
 DB 683 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 741
 QY 767 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 824
 DB 742 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 801
 QY 825 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 884
 DB 802 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 861
 QY 885 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 944
 DB 862 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 921
 QY 945 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1004
 DB 922 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 981
 QY 1005 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1064
 DB 982 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1041
 QY 1065 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1123
 DB 1042 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1096
 QY 1124 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1183
 DB 1097 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1156
 QY 1184 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1242
 DB 1157 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1216
 QY 1243 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1302
 DB 1217 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1276
 QY 1303 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1336
 DB 1277 LITNKSVPVRYVGMFTDDGQYMAHKTINYEIETLLKARIKYVSGGQWRNQOVNSE 1310

RESULT 6

GTF2_STRDO

ID GTF2_STRDO STANDARD; PRT; 1592 AA.

AC P27470;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 13-SEP-2005 (Rel. 48, Last annotation update)

DE Glucosyltransferase-I precursor (BC 2.4.1.5) (GTF-I) (Dextranucrase)

DE (Sucrose 6-glucosyltransferase).

OS Streptococcus downei (Streptococcus sobrinus).

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1317;

RN [1]

RESULT 7
 GTFD_STRMU
 ID GTFD_STRMU STANDARD; PRT; 1462 AA.
 AC P49331; O69383; O69386; O69389; O69392; O69398;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranase)
 DE (Sucrose 6-glucosyltransferase).
 GN Name=gTfD; OrderedLocusNames=SMU_910;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RX MEDLINE=91100958; PubMed=2148600;
 RA Honda O., Kato C., Kuramitsu H.K.;
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
 the glucosyltransferase-S enzyme."
 RL J. Gen. Microbiol. 136:2099-2105(1990).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
 RC MT4467 / Serotype e, and MT8148 / Serotype c;
 RX MEDLINE=98231643; PubMed=9570124;
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
 RA Kimura S., Hamada S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 Streptococcus mutans."
 RL FEMS Microbiol. Lett. 161:331-336(1998).
 RN [3]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=UAI159 / ATCC 700610 / Serotype c;
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
 pathogen."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to
 play a key role in the development of the dental plaque because of
 their ability to adhere to smooth surfaces and mediate the
 aggregation of bacterial cells and food debris.
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
 fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
 water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
 forms of glucans.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -!- SIMILARITY: Contains 6 cell wall binding repeats.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use as long as its content is in no way modified and this statement is not
 removed.
 CC -----
 CC EMBL; M29296; AAA26895.1; -; Genomic DNA.
 CC EMBL; D88653; BAA26103.1; -; Genomic DNA.
 CC EMBL; D88656; BAA26107.1; -; Genomic DNA.
 CC EMBL; D88659; BAA26111.1; -; Genomic DNA.
 CC EMBL; D88662; BAA26115.1; -; Genomic DNA.
 CC EMBL; D89979; BAA26121.1; -; Genomic DNA.
 CC EMBL; AE014932; AAN58619.1; -; Genomic DNA.
 CC HSSP; P06653; 1GYM.
 DR InterPro; IPR002479; Cell_wall_bd_put.

DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 4.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
 KW Transferrase.
 FT SIGNAL 1 ? Potential.
 FT CHAIN ? Glucosyltransferase-S.
 FT REPEAT 1232 1295 1.
 FT REPEAT 1296 1359 2.
 FT REPEAT 1360 1423 3.
 FT REGION 1232 1423 3 X 63 AA approximate tandem repeats.
 FT VARIANT 10 10 Y -> H (in strain GS-5, strain MT4239,
 strain MT4245, strain MT4251, strain
 MT4467 and strain MT8148).
 FT VARIANT 19 19 I -> V (in strain GS-5, strain MT4239,
 strain MT4245, strain MT4251, strain
 MT4467 and strain MT8148).
 FT VARIANT 58 58 K -> E (in strain MT4467).
 FT VARIANT 68 68 A -> S (in strain MT4239 and strain
 MT4245).
 FT VARIANT 81 81 A -> T (in strain MT4251 and strain
 MT8148).
 FT VARIANT 113 113 T -> I (in strain MT4239 and strain
 MT4245).
 FT VARIANT 122 122 A -> V (in strain MT4239, strain MT4245
 and strain MT8148).
 FT VARIANT 132 132 A -> S (in strain GS-5 and strain
 MT4467).
 FT VARIANT 135 135 A -> V (in strain MT4245).
 FT VARIANT 137 137 A -> T (in strain GS-5, strain MT4239,
 strain MT4245, strain MT4251, strain
 MT4467 and strain MT8148).
 FT VARIANT 202 202 V -> L (in strain MT4239).
 FT VARIANT 255 255 E -> D (in strain MT4239, strain MT4245
 and strain MT4251).
 FT VARIANT 275 275 D -> N (in strain MT4239, strain MT4245
 and strain MT4251).
 FT VARIANT 288 288 D -> N (in strain MT4239, strain MT4245
 and strain MT4251).
 FT VARIANT 301 301 Q -> H (in strain MT4245).
 FT VARIANT 313 313 D -> N (in strain MT4239 and strain
 MT4251).
 FT VARIANT 317 317 E -> K (in strain MT4239).
 FT VARIANT 328 328 F -> F (in strain MT4239).
 FT VARIANT 350 350 F -> L (in strain MT4239, strain MT4251
 and strain MT4467).
 FT VARIANT 633 633 KKQYIQ -> EKEVTL (in strain MT4251).
 FT VARIANT 688 688 A -> S (in strain MT4239).
 FT VARIANT 726 732 TDQSEA -> ADKGNDS (in strain MT4251).
 FT VARIANT 726 730 TDQGS -> ADKGN (in strain MT4239 and
 strain MT4245).
 FT VARIANT 762 762 T -> A (in strain GS-5, strain MT4239,
 strain MT4245, strain MT4251, strain
 MT4467 and strain MT8148).
 FT VARIANT 964 964 D -> Y (in strain MT4251).
 FT VARIANT 1019 1019 E -> K (in strain MT4245 and strain
 MT4251).
 FT VARIANT 1059 1060 LG -> IR (in strain MT4251).
 FT VARIANT 1060 1060 G -> R (in strain MT4245).
 FT VARIANT 1080 1080 G -> R (in strain MT4239).
 FT VARIANT 1142 1142 H -> Q (in strain GS-5).
 FT VARIANT 1198 1198 S -> N (in strain MT4239).
 FT VARIANT 1220 1220 Y -> C (in strain MT4251 and strain
 MT4467).
 FT VARIANT 1280 1280 F -> L (in strain MT4467).
 FT VARIANT 1282 1282 Q -> P (in strain MT4245).
 FT VARIANT 1290 1290 K -> T (in strain MT4245).
 FT VARIANT 1311 1311 G -> D (in strain MT4245).
 FT VARIANT 1403 1403 G -> D (in strain GS-5 and strain
 MT4467).
 FT VARIANT 1425 1425 R -> K (in strain GS-5).
 FT VARIANT 1449 1449 R -> K (in strain MT4467).
 FT VARIANT 1428 1462 RYDKNSGMVNVKVTLANGRRIGIDRWGIARY -> VY
 R (in Ref. 1).

SQ	SEQUENCE	1462 AA; 163388 MW; CE4A279C4D708645 CRC64;
	Query Match	50.1%; Score 3579.5; DB 1; Length 1462;
	Best Local Similarity	51.9%; Pred. No. 3e-162;
	Matches	711; Conservative 220; Mismatches 375; Indels 63; Gaps 20;
QY	1	MEKKYRFLKRVKRWVTSIASAVVTL--TSLGSLVLRKADSTDDRQQAFTESQASLVTTTS 59
DB	1	METKRYKMYKKGHWITAVASGLITLGTTLTSGSVSAB-----TEQQTSDKVV 51
QY	60	EAAKETLTATDSTATSQPTATVTDNVSTTNOSTNTANTANFVVKPTT---SEQA 116
DB	52	QKSEDDKAASE-----SSQTDAPTKQAQTEQQAQSQANVADTSIYKETPSQNIITQA 107
QY	117	KTDSNDKIITTSKAVNRLTATKFPVPAANNHTAHPKVTVDKIPIKPKIKLQPSLSQD 176
DB	108	NSD--DKVTNTKSEBAQSEERTQAEB--AQTASSQALTOAKAELTKQRQTAAQENK 163
QY	177	DTAALGNVKNIRKVGKYYKEDGTLQKYALNINKGTFFPDE--TGALSNNLTLSKKN 235
DB	164	NPVDLAAIPNVKIDPKYIIGSDGQPKKFPALTNNKVLFPDKNTGALTDSQYQKQ 223
QY	236	IT--NNDNTNSFAQNOVYSTDVANFHDVHYLTAESWYRPKYILKDGKTWTQSTEKDPR 293
DB	224	LTFLAND----YTPHQIVNPENTSELETIDNVITADSWYRPKDIILKNGKTKWTASSEDLR 279
QY	294	PLMTWPDQETQROQVYNNMAQ--LGIHQTYNTATSPLOLNLAQTIQTKIEKITAENK 352
DB	280	PLLSWMPDKQTOIAYLNNMQGLGTGENYTADSSQESLNLAQTVQVKIETKISQTOQ 339
QY	353	TNWLQRTISAFVKTOSAMNSDEKFPD-----DHLQKGLLYSNNSKLTQSQANSYRLNR 408
DB	340	TQWLRIIINSFVKTPFNWNSQTESDAGEKDHQGLLYSNSDK--TAYANSRLLNLR 398
QY	409	TPTNQTKKDPYTRADRTIGGYEFLANDVNSNPVQAEQLNLHLFMNFCNIVANDPD 468
DB	399	TTSQTGK--PKYFEDNSGGYDFLLANDIDNSNPVQAEQLNLHLFMNYSIIVANDPE 456
QY	469	ANFDSIRVDADVNDADLLQIAGDYLKAAKGIIHKNDKAANDHLSILEAWSYNDTPYLHDD 528
DB	457	ANFDGVRVADVNVNADLLQIASDYLKAHYGVKSEKNAIHLNLSILEAWSNDNDPQVNTKT 516
QY	529	GDNMINMDNRLSLILYSLAKPLNQ-----RSGNPLIITNSLVNRTDNDNAETAAPS 580
DB	517	KGAQLPIDNKLRLSLYALTREPLEKDNASKNIRSGLEFVITNSLNNRSAEKGNSERMAN 576
QY	581	YGFIRAHDEVDLIRNIIRTEINPNVGYSTTBEBIKAFIYNKDLATEKKYTHYNT 640
DB	577	YIFIRAHDESVTVAKIILKAQINPKTDGLTTLDELKQAFKIYNEDMRQAKKYTQSN 636
QY	641	ALSAYALLTNKSVPRVYVYDMDTDDGQYMAHKTINYEAIETLLKARIKYVSGGQAMRNQ 700
DB	637	PTAYALMLNSKDSITRLYVYDGYSDGQYMATKSPYDAIDTLKARIKYAAGGQDMKIT 696
QY	701	QVGNSE-----IITSVRVKGKALKATDGTDRTRTSGVAVIEGNPNPSRLKASDRV 751
DB	697	YVEGDKSHMDWDYTGTLTSVRYGTGANEATDQGEATKTQGMAVITSNNSPSLKNQNDKV 756
QY	752	VNNMGAHNQAVRPLLLTTDNGIKAYHSDQEAAGLVRVTNDRGELIFTAADIKGYANPQ 811
DB	757	IYVMGTAKNQBYRELLLTLDKGLTSYSDAAKSLYRKTNKGELVFDASDIQGLYNPQ 816
QY	812	VSGYLGVWVPVGAADQDVRVAASTAPSTDGKS VHQNALDSRVMPFGFSNFPQAFATKBE 871
DB	817	VSGYLAVWVPVGASDNQDVRVAASNKANATQVYESSSALDSQLIVEGFSNFDFTKDS 876
QY	872	EYTNVVIANKVDKFAEWGTYTDEMAYQYVSSYTDGSLDSVIQNGYAFTRDYDLIGSKPNK 931
DB	877	DYTNKKIAQNVQLFKSWGTYTSEMAYQYVSSYDGSFLDSIIQNGYAFEDRYDLAWSKNNK 936
QY	932	YGTADDLVKAIKALHSGIKVWADWVPDQMYALPEKEVVTATRVKDYGPVAGSOLKNTL 991
DB	937	YGSQDMINAVALLHSGIQLVITADWVPDQIYNLPKEVVVATRVNDYGEYRDKSELKNTL 996

QY	992	YVVDKSSGKQQAQYKGAFLBELOAQYPELPARKQISTGVPMDPSPVKIKQWSAKYFNCT 1051
DB	997	YAANTKSKNGKQYQAKYGGAFLSLAAPYSIFNRTQISNGKIKIDPSEKITAWKAKYFNCT 1056
QY	1052	NILGRGAGYVLKQDQATNTYFSLVSDNTFLPKSLVNPNGHTSSVTGLVFDGKGVVYVTS 1111
DB	1057	NILGRGVYVLKDNASDKYFELKGNQTYLPQOMTY-----KEASTGFVNDGNGMTFYSTS 1111
QY	1112	GNQAKNAFI--SIGNNWYFDNNGYVMTGAQSIINGANYFSLNSGIQLRNAIYDNGNKVLSY 1170
DB	1112	GVQAKNSFVQDAKGNWYFDNNGHMYGLQHLNGEVQYFSLNGVQLRESFLENADGSKY 1171
QY	1171	YGNDRRYVNGYFLP--CQWRYFO--NGIMAVGLTRVHGAVOYFPDASGFQAKGQBITTAD 1227
DB	1172	FCHLGNRYSGYSPDNDKRWYFPDASGVMAVGLKTINGNTQYFPDQDGYQVKGAWITGSD 1231
QY	1228	GKLYRFDSDSGNISNRFVRNSKGEWFLFDHNGVAVTGTVTFNGORLYFKPNGVQAKGEF 1287
DB	1232	GKRYFDDGSGNMAVRPANDKNGDWYLYNSDIALVGQTINGKTYIFGQDGKQIKGKI 1291
QY	1288	IRDANGYLRYDPPNSGNEVRNRFVRNSKGEWFLFDHNGIAGVATGARVVANG 1336
DB	1292	ITD--NGKLYFLANSSELARNIFATDSQNNWYFGSDGVAVTGSQTIAG 1339
RESULT 8		
Q54178	STRGN	
ID	Q54178_STRGN PRELIMINARY; PRT; 1577 AA.	
AC	Q54178_Q54247;	
DT	01-NOV-1996 (TRENBLrel. 01, Created)	
DT	01-NOV-1996 (TRENBLrel. 01, Last sequence update)	
DT	01-JUN-2003 (TRENBLrel. 24, Last annotation update)	
DE	Glucosyltransferase.	
GN	Name=gtfG;	
OS	Streptococcus gordonii.	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OC	Streptococcus.	
OX	NCBI_TaxID=1302;	
RN	(1)	
RP	NUCLEOTIDE SEQUENCE.	
RC	STRAIN=Challis;	
EX	MEDLINE=96157084; PubMed=8586195;	
RA	Vickerman M.M., Sulavik M.C., Clewell D.B.;	
RT	"Molecular analysis of Streptococcus gordonii glucosyltransferase phase variants.";	
RL	Dev. Biol. Stand. 85:309-314 (1995).	
RP	(2)	
RC	NUCLEOTIDE SEQUENCE.	
RA	Minick P., Vickerman M.;	
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; U12643; AAC3483.1; -; Genomic_DNA.	
DR	PIR; B41898; B41898.	
DR	HSP; P06653; 1HCX.	
DR	GO; GO:0016740; F:transferase activity; IEA.	
DR	GO; GO:0009250; P:glucan biosynthesis; IEA.	
DR	InterPro; IPR002479; CW binding.	
DR	InterPro; IPR003318; Glyco_hydro_70.	
DR	Pfam; PF01473; CW binding_1; 5.	
DR	Pfam; PF02324; Glyco_hydro_70; 1.	
KW	Transferase.	
SQ	SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;	
Query Match		
Best Local Similarity 49.6%; Pred. No. 5e-152;		
Matches 695; Conservative 224; Mismatches 398; Indels 85; Gaps 26;		
QY	1	MEKKYRFLKRVKRWVTSIASAVVTLT---SLGSLVLRKADS---TDDRQQAFTESQ 52
DB	2	MEKKVHYKMKHKVKNWAIAVTSLALLVAPKALGESGVYADDANQVTNVKEQSAVQSK 61
QY	53	ASLVTTSEAA-----KETLTATDSTATSQPTATVTDNVSTTNQ-----SINTTA 100

Db 62 DSEQTSDKATDSSQLEVEQOASSSKETQYQASATNPTA-----NEQTQOQDEKETSRTD 117
QY 101 NTANFVVKPT-TTSEQATDNGDKIITTSKANRLTA-----TGKRVFVANNNTAHP 150
Db 118 SRHELTKTSDSSEKSGSQEPKQVADQAESTDQTAQALQAKQDSRANDQEBETTENA-K 176
QY 151 KTVTDKIVIPKPIKGIKGLKOPS-----SLSQDDIAALGN-----VKNIRKVNKGKY 194
Db 177 ATVSDKII-ATPKERLPEPAQKESITEKMLAAQAQAAPVTEHDDDDVLAHIKTIIDGKK 235
QY 195 YYYKEDGTLQKYNALNINGKTFDD-ETGAL-SNNTLPSKGNINNDNTNSFAQVNOVY 252
Db 236 YTVQDGTGKKNFAVELNGKILYFADTAETGALVDSNEYQFQOQTSSLN---NEFTQKNAFY 292
QY 253 STDVANFEHVDHYLTAESWYRKYILKDGKTTQSTQKDFRPLMTWMPDQETQRYVNY 312
Db 293 GTTDXDIEVDGYLTADSWYRKFILKOGKTWTASTETDLRPLLMANWPKRTQINLYNY 352
QY 313 MNAQ-LGIHQTYNTATSPQLNLAAQTIQTKIEEKITAEKNTNWLRTQISAFVKQTQSAWN 371
Db 353 MNQENLGI-GAPESKTEQVLLTNVAVQVQRKIEERISKEGDTKMLRTLMSAFVKTPQNN 411
QY 372 -----SDSEKPPDDHLQKGLLYSNNSKLTQANSYRILNRPTTQTKGKDPRTADRTI 427
Db 412 IKTESETTGTNRKHLQGGALLYTNSDK-TSHANSRYRLNRPTTQTKG--TPKYFIDKSN 468
QY 428 GGYEFLLANDVNSNPVQAEQLNLWHLPLMNGNIYANDPDANFDSIRVDAVDNVDADLL 487
Db 469 GGYEFLLANDFNSNPAQAEQLNLWHLHYMNGFSGIVANDPTANFDGVRVDAVDNVDADLL 528
QY 488 QIAGDYLKAAKGIHKNDAANDHLSILEAWSYNDTPYHLDDGDNDNMNDRRLSLYLXSL 547
Db 529 QIADYFKSRYKVGSEEEALKHLSILEAWSNDPDYKNTKGAQLAIDNKLRLSLLYSF 588
QY 548 AKPLNQRGMPLIINSLVNRDNDNAETAAPSYSFIRAHDSVQDLIRNIIRTEINPNV 607
Db 589 MRKLSIRSGVEPTIINSLNDRSTENKNGERTANYIFVRAHDSVQTVIADIIRENINPT 648
QY 608 VGYSFTTEIKKAFIYKOLLATKTKYTHYNTALSYALLTNKSSVRVYVYGDFTDGG 667
Db 649 DGLTTEMDLQKAFKINEDMRKADKKYQFNIPTAHALMLSNKDSITRVYGDLYTDDG 708
QY 668 QYMAHTINYEAIETLLKARIKYVSGGQAMRQVQ-----NSEIITSVRVYKGGAL 718
Db 709 QYMEKKSPTHDAIDALLARIKYVAGGQDMKYTVNGVPREADKWSYNGILTSVRVGTGN 768
QY 719 KATDGTDRTRTSGVAVIEGNPSRLKASDRVVVNMGAHKNQAYRPLLLTDDNGIKAY 778
Db 769 EATDGTGTAETRTQGMVAVIASNPNLKLNEWDKLVNMGAAHKNQYRPLVLLTTKDGISRY 828
QY 779 HSDQEA-AGLVRYTNDRGELIFTAADIKGYANPOVSGYLGWVVPVGAADQDVRVAATA 837
Db 829 LTDEEVPQSLMKTKDANGILTDFMDNDIAGYSNVQVSGYLAVVVPVGAKENQDARVTSKK 888
QY 838 PSTDCKSVHQAALDSRVWFEGSFNFQAFATKKEBYTNVIAKQVYDKEAEGVTFDEMAP 897
Db 889 KNASQVYVESPALDSQLIYEGFSNFQDFATDDQYTNKVIKAVNLKAEWGVTSFELPP 948
QY 898 QYVSTGDSFLDSVTQNGYAFDTRDYLGISKPNKYGTADDLVKAIKALHSGIKVYMAOW 957
Db 949 QYVSDQGTFLDSIIQNGYAFEDRYDMAMSKNNKYGSLDILLNLRALHLSVNIQAIADW 1008
QY 958 PQQVYALPEKEVWTRVDYKGTVPVAGSQIKNTLYVDGKSSGQDQQAQYGGAFLEBLOA 1017
Db 1009 PDQIYNLPKGEVWTRVNNYGTREYGAIEIKELVYANSKTINGDYQYQYGGAFLEBLOA 1068
QY 1018 KYPELFARQKISTGVPMPSVKIKOWSAKYFNGTNILGRGAGYVLKDAQNTYFSLVSDN 1077
Db 1069 KYPELFARVQISNGQKMTTDEKIKWSAKYFNGTNILGRGAGYVYVLDKMGSEKYLKNGE 1128
QY 1078 TFLPKSLVNPNGHTSSSYTGLVFDGKGYVYVYSTSGNQAKNAFISLGN-NWYFFDNNGYV 1136

Db 1129 TALPQKLVN-----KEASTGFVKDTNGFKFYSTSGNAQKDTFIQDENGNNWYFDNQGLV 1183
QY 1137 TCAQOSINGANYFSLNGIOLRNAIYDNGKNKLSYTGNDGRRYRNGYVYLF-GQOMRYPO-N 1194
Db 1184 TGAREIDGKQLYFMKNGVQLRDALQEDENGNGYQYDKTGAQVLRNRYTSDGQNWRYFDAK 1243
QY 1195 GIMAVGLTRVHGAVQYFDASGPOAKGQFITTTADGKLYRFDKSGNQSISRFRV-NKSGEW 1253
Db 1244 GYMARGLVKIGDQYFDQNGYQVKGKVRADKGLRYFDKSGNAVINRFAQGDNPDSW 1303
QY 1254 FLFDHNGVAVTGTVTFGQRLYFKPENGVOAKGEFIRDANGLYRYDPSNGNEVRNFRVN 1313
Db 1304 YTFGADGVALTQLQKIGQOQTLFYGDGKQVQKQVQVMLADKSIYFYDANSGEVAVNFAEG 1363
QY 1314 SKGEWFLPDHNGIATGARVNN 1335
Db 1364 AKNEWYFDDQKAVTGLKTIIN 1385

RESULT 9

Q69A94 LEUME
ID Q69A94 LEUME PRELIMINARY; PRT; 1454 AA.
AC Q69A94;
DT 25-OCT-2004 (TremBLrel. 28, Created)
DT 25-OCT-2004 (TremBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TremBLrel. 28, Last annotation update)
DE Dextranucrase (EC 2.4.1.5).
GN Name-darP;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IBT-PQ;
RA Fernandez-Vazquez J.L., Lopez-Munguia A., Olivera C.;
RT "Molecular characterization of a dextranucrase gene from Leuconostoc
mesenteroides IBT-PQ isolated from pulque."
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY504865; AAS79426.1; -; Genomic DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0004675; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; F:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1454 AA; 160569 MW; 2DC734296386FD87 CRC64;

Query Match 46.9%; Score 3348.5; DB 2; Length 1454;
Best Local Similarity 51.6%; Pred. No. 3.2e-151;
Matches 699; Conservative 195; Mismatches 392; Indels 69; Gaps 27;
QY 17 VTVSIASAVVTLTSLGSLVKADSTDDRQQAQVTSQASLVTTSEAAKETLTATDTSTATS 76
Db 32 VTSFIQQAQADVQSKNGVVVIT-----AVQNSDATTTD-----KSIITNDKATTTA 79
QY 77 ATS-OPHTATVDNVSTTQSTNTTANTANFVVKPTTSEQAKTDSNDKLIITTSKAVNRLT 135
Db 80 DTSTNDKATTTADTSTNDKATTTADTSTN--DKATTTAD--ISTNNKATTTADTSTNNKA 135
QY 136 ATGFPVPAANNHTAHPKTVTDKIVPIKPIKIGKIKQPSLSQDDIAAL-GNVKNIRKVNKGKY 194
Db 136 ATTADTSDNNNSA---TTSDKDVSLAOKSAIDNNSKTITDPAASFEASKMLKLTIDGT 192
QY 195 YYYKEDGTLQKYNALNINGKTFDD-ETGALSN-NLTPSKGNINNDNTNSFAQVNOVY 252
Db 193 YYYDNGQIKNEATVIDGKLYFDKDTGALADTNDYQFLEGLTSEN---NNYTEHNASV 249
QY 253 STDVANFEHVDHYLTAESWYRKYILKDGKTTQSTQKDFRPLMTWMPDQETQRYVNY 312
Db 250 GTSSASVNTVDGYLTADSWYRKFILKOGKTWTASTETDLRPLLMANWPKRTQINLYNY 306


```
QY 479 VDNVADLLQIAGDYLKAAKGIHKNDKAAANDHLSILEAWSYNDTPYLLHDDGDMNINMDNR 538
Db 518 VDNVADLLQIAGDYPKSYKVGSEEEAIKHLISLEAWSDNDPDYKNTKGAQLAIDNK 577
QY 539 LRLSLLYSLAKPLNQSGMPLTNSLVNRTDNDNAETAAVPSPSFRADHSEVQDILRNI 598
Db 578 LRLSLLYSFWNLISRSVEPTITNSLNDSSSEKNGERMANIYFVRADHSEVQTVIADI 637
QY 599 IRTEINPNVGVSTTEBIKAFPEIYNKOLLATEKKYTHYNTALSALLTNKSSVPRVY 658
Db 638 IRENINPNVDTGLTFTMDLQAKFIYNEDMRKADKKYQFNIPTAHALMSLNKSDITRVY 697
QY 659 YGDMPTDDGOYMAHKTIYNEAETLLKARIKYVSGGAMRQVQV-----NSEIIT 709
Db 698 YGDLTYDDGOYMEKKSPYDAIDALLRARIKYVAGQDMKYVMYGPREADKWSYNGILT 757
QY 710 SVRYGKALKATDGTDRTRTSVAVTEGNPNLSRLKASDRVVVNMGAHKNQAYRPLL 769
Db 758 SVRYGTGANEATDEGTAEFTQGMVAVIASNNPNLKNLNEWDLQVNMGAHKNQAYRPLL 817
QY 770 TTONGIKAYHSDQEA-AGLVRYTNRDELIFTAADIKYANPQVSGYLVVYVPGAAADQ 828
Db 818 TTQKDGISRYLTDEVPQSLWKKTDANGILTFDMNDIAGYSNVQVSGYLAWVYVPGAKADQ 877
QY 829 DIVEVAASTAPSTDCKSVHQNALDSRVPFEGSNFQAPATKKEEYTNVVIKNDVKPAEW 888
Db 878 DARTTAKSKKNASGVYESSAALDSQLIYEGFSNFQDFATRDDQYTNKVIKNNVLFKEW 937
QY 889 GVTDPMAPOYVSSYSDGSLDSVIQNGYAFTRDYDIGISPKNYGTTADLLVAKALHSK 948
Db 938 GVTSELPPOYVSSQDGTFLDSITQNGYAFEDRYDMASKNNKYGSLKOLLNLRALHSV 997
QY 949 GIKVMADWPQOMYALPEKEVVTATRVKDYGPVAGSQIKNTLYVVDGSSGKDDQQAQYK 1008
Db 998 NIQATADWPQOYLPLGKEVVTATRVNNGYTYREGAEIKEKLYVANSKTNETDFQCKYK 1057
QY 1009 GAFLBELQAKYFELPARKOISTGVPMDPVSKIKQWSAKYFNGTNIILGRGAGYVLKQOATN 1068
Db 1058 GAFLDELKAKYFELPERVOISNGQKMTTDEKITKWSAKYFNGTNIILGRGAGYVLKQWASN 1117
QY 1069 TYFSLVSDNTELPKSLVNPNGHTSSSVTGLVFDGKGYVYVSTGNOAKNAPISLGN-WKY 1127
Db 1118 DYLTRNGEIVLPKLVNKN-----SYTGFSNDANGTFKYSTGYQAKNSFIDENGWY 1172
QY 1128 YFDNNGYMYTGAQSGINGANYFSLNGIQLRNALYDNGNKVLSYGYNDGRRYENGYYLF-G 1186
Db 1173 YFDKRGYLVGTGAHEIDGKHVYFLKNGIQLRDSIREDENGQYYYDQTGAQVLRYYTDTG 1232
QY 1187 QQWRYFO-NGIMAVGLTRVHGAQVFPDASGFOAKQFIITADGKLYFRDRSGNOISNRF 1245
Db 1233 QWRYFDKAGVNMARGLVKIGDQOQFPDENGQYQVKGIKVSADGKLYRFPKDSGNVIRNF 1292
QY 1246 VR-NSKGWFLPDHNGVATGVTGNGQRLYKFPNGVQAKGEFFIRDANGYLYBYDPNSNG 1304
Db 1293 AGDNPDSWYFVGVFAKLTLGQKTLGQOQLYFDQDGQVKGIVTLSDSKSIRYFANSGE 1352
QY 1305 EYVRNFRVNSKGEWFLPDHNGIATGARVV 1334
Db 1353 MAVGFAEAGKNEWYFDKTAIVGLQKI 1382
```

RESULT 11

```
Q55265 STRSL
ID Q55265 STRSL PRELIMINARY; PRT; 1577 AA.
AC Q55265;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Name=gtfm;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
```

```
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL; L35928; AAC41413.1; -; Genomic_DNA.
DR PIR; T30858; T30858.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR004829; C:surface antigen.
DR InterPro; IPR002479; C:surface antigen.
DR InterPro; IPR003318; G:glycohydro_70.
DR Pfam; PF01473; C:binding_1; 4.
DR Pfam; PF02324; Glycohydro_70; 1.
DR ProDom; PD153432; C:surface_antigen; 1.
DR Signal; Transferase.
KW SIGNAL
FT CHAIN 1 38 Potential.
FT SIGNAL 39 1577 Glucosyltransferase.
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7BF3 CRC64;
```

Query Match 45.0%; Score 3216; DB 2; Length 1577;

Best Local Similarity 47.7%; Pred. No. 7.7e-145;

Matches 687; Conservative 214; Mismatches 413; Indels 126; Gaps 33;

QY 1 MEKVRPKRUKRUKRUVTVSIAAVVTLTSLG-----SLVKADST-----DDRQAVT 49

Db 1 MENKVRPKRUKRUKRUVTVT--LSMVALAGSLLAQKGVADSTSAFNGDGLQQLSE 58

QY 50 ESQASLVTS-----EAAKETLTATDTST-----ATSATSO-----PAT-- 84

Db 59 DGTASLVTTTTEQASQASVSATASVSHTSFQAATSQATSAQATQATQSPVASQE 118

QY 85 ---VDNVSTNTSTNTT-----ANTANFVVKPT-----TTSQAK---TDNSDKLITTS 128

Db 119 VAVSSQTQSSQGETTTEQVSGQISTQVAGTSAQSTPSVTQAPRVLTAAPALATR 178

QY 129 KA-----VNRLTATGKFPANNHTAHPKTVTKIVPIKPI----- 164

Db 179 AADSTIRINANNRNTITITASGTTTNPVTITGENTP-KPNVTVTSNTPNVTITVTPN 237

QY 165 --GKLKPSLSQDDIAALGN-----VKNIRKVGKYYKYKEDGLTKNYALNING 213

Db 238 QNKVPQSPQSPQPKNPQPNQPSLDYKPVASNLKTIDGKQYV-ENGVVKNAALELDG 296

QY 214 KTFEFDGALSNNTLPSKKGN-ITNNDNTNSFAQYNQVYSTDVANFEHVDHYLTAESWY 272

Db 297 RLYVFEDEGTAMVDQSKPLYRAADAI PNN---SIYAVYNOAYDTSSKSEFHLDNFLTADSWY 353

QY 273 RPKYILKDGKTWTQSTEKDFRPLMTWPDQSTQRYVNMNAQLGIHGYNTATSPLOL 332

Db 354 RPKQLIDGKKTWASTEKDYRPLMTWPDQSTQRYVNMNAQLGIHGYNTATSPLOL 413

QY 333 NLAQOTIOTKTEEKITAENKTNMLRQTISA FVKTSQANSDSEKPP---DDHLOKALLY 389

Db 414 AAAAETVORGEEERIGREGNTTTLRLQLMSDFIKTOPGMNSSEEDNLLVGKDLHQQGALT 473

QY 390 SNNSKLTQANSNYRILNRTPTNQTGKDPRTYADRTIGGYEFLFLANDVNSNPVQAEQ 449

Db 474 LNNNS-ATSHANSDFRLMRTPTNQTGTR--KTHIDRSNGGYELLANDIDNSNPVQAEQ 530

QY 450 LNLWHLFANFGNIYANDPDANFDSIRVDVNDVADLLOIAGDYKAAKGIHKNDKAND 509

Db 531 LNLWHLFANFGNIYANDPDANFDSIRVDVNDVADLLOIAGDYKAAKGIHKNDKAND 590

QY 510 HLSILEAWSYNDTPYLLHDDGDMNINMDNRRLSLYSLAKPLNQSGMPLTNSLVNRT 569

Db 591 HLSILEAWSYNDTPYLLHDDGDMNINMDNRRLSLYSLAKPLNQSGMPLTNSLVNRT 650

QY 570 DDNAETAAPVSPYSFTRAHDSEVQDLIRNIIRTEINPNVGVSTTEEIKKAFEYINKDLL 629


```
Db      866  VQDAQYTNKKIAENTDLFKAWGVTSEFAPQVVSATDGTFLDSIIQNGYAFSDRYDLAM 925
QY      927  SKNKYGTADDLKAIALKALHSGIKVMADWVPDQYALPEKEVVTATRDVKYCTPVAGSQ 986
Db      926  SKNKYGSKEDLANALKAALHAGIQAIADWVPDQYIQLPEKEVVTASRDVNYGRVKIDQP 985
QY      987  IKNTLVVDPKSGKDDQQAQYGGAFLEELQAKYPELFARKQISTGVPMDPSPVKIKOWSAK 1046
Db      986  MVNKLVLANTKSSGKDFQAKYGGCEFLAELQKQYPEMFTAKMISTGKPIDSSVKLKWSAQ 1045
QY      1047  YFNGTNILGRCAGYVLKQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGLVDFDGKGV 1106
Db      1046  YFNGTNVLGRGTDYVLSDEGTGKYFTVNEKGEFLPAVLT----GDKEAKTGFYNDGKGMT 1101
QY      1107  YYSTSGNQAKAFISLGNWYFDDNNGYMTGAQSLNGANYFLPSLNGIOLRAIY--DNG 1164
Db      1102  YFTTAGSQAQSPVTVAGNTYFYDYGHWVTGNGTINTFYFLPPLNGVLMKDAVMBDDR 1161
QY      1165  NKVLSYYGNDGRYE----NGYLF----GQ-QMRYFQON-GIMAVGLTRVHGAVQYFDAS 1214
Db      1162  RSV--YYGKTGVNKGSRNWFAMTDSKQLRFRHFDNYGFMVGLVTHGNVQYIDEE 1219
QY      1215  GFQAKQFITTADGKLRIFDRSGNQISNRFRVNSKGWFLPDHNGVAVTGTVTNGQRL 1274
Db      1220  GFQVKEBDFYTDKAGQTRYFDKNTGNLVKGQFF-NQNGHWYSSDDQLIAKGAQTIKGQKL 1278
QY      1275  YF-KPENGVAQGEFIRDANGLYRYDPNSGNEVRNFRVNSKGWFLPDHNGIYAVTGARV 1333
Db      1279  YFDATGAQVKGDFVTDKDGNTFFYSGDGTGLAVSTFFSTGNNAMFYADENGHVAKGEKT 1338
QY      1334  VNG 1336
Db      1339  ING 1341

RESULT 13
ID Q56CX8_9STRE PRELIMINARY; PRT; 1506 AA.
AC Q56CX8.
DT 10-MAY-2005 (TReMBLrel. 30, Created)
DT 10-MAY-2005 (TReMBLrel. 30, Last sequence update)
DE Glucosyltransferase-t.
GN Name=glft;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY966490; AAX76986.1; -, Genomic_DNA.
KW Transferase.
SQ SEQUENCE 1506 AA; 167016 MW; 94B882EF2C17C451 CRC64;

Query Match 44.5%; Score 3175.5; DB 2; Length 1506;
Best Local Similarity 46.6%; Pred. No. 6.2e-143;
Matches 645; Conservative 237; Mismatches 404; Indels 99; Gaps 25;

QY      1 MEKKYRPLKRYKKEWTVSTASAVVLTLSLGSVLKADSTDD--RQAVTESQASLVT 58
Db      1 MERKLHKLHKVQKWQVTVIASAGLASIVGAGLSQTVSADLLAKEQAAASQQKAAANQ 60
QY      59 SE-----AAKETLTATDTSATSQPTATVTDNVSTNTSTNTTANTANFVVK----- 108
Db      61 NEDEVASDAADTASAKATSEKEVQSSDTSNSETNQVETKQQA--SAKESADAVAQAQA 118
QY      109 -PTTTSQAKTDSKIITTSKAVNRLTATGKVPANNNTAHPKTVTDKIVPKPKIGKL 167
Db      119 GPATTSQVASSSESS--VAPSKADKAAA----- 145
```

```
QY      168  KQPSSLSODDIAALGNVKNIRKNGKYVYKEDGTLLQKNVALNINGKTPFD--ETGALSN 226
Db      146  ---GSVQNEEERAAISLANIKKIDGKYVTMADGSKGNFAITVDCQMLFDARTGALSS 202
QY      227  NTLPSKKGNTNDNTNSPAQNVYSTDVANFEHYDHYLTAEBSWYRPKYILKDGKTWTQ 286
Db      203  TSYTSPSQGLT--PIVSDFSVNNAKAFDSSEKSPFLVDGYLTAEBSWYRPKILENGKTWVD 260
QY      287  STEKOPRPLMTWPDQETORQYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 346
Db      261  SKETDLRPVLMGWPKNQTVAYLNTMSKALGKEEFTTETSQLTTLTAELIAQKIEAR 320
QY      347  ITAEKNTNLRQITISAFVKTQSAWNSDSEKPPD--DHLQKGALLYNNNSKLTQSAANSYR 404
Db      321  VSEKQGTWLRAMAFVATQSRWNKDSQ--YDKADHLOGGALLYNNNN--LTEWANSNR 378
QY      405  ILNRTPTNQTGKDPRTYATRTIGGYEFLANDVNSNPVQAEQNLWLHFLNFGNIYA 464
Db      379  LLNRTPTRGDK--THYSKADKYGGYEFLLANDVNSNPVQAEMLNQIHYLMNWGEIVM 436
QY      465  NDPDANFDSIRVDVNDVADLLOIAGDYLKAAGKGIHKNDKAANDHLSILEANSYNDTPY 524
Db      437  GDKANFDPGIRVDVNDVNADTLQLYTNYFNSVYGVNKSQAQALAHISVLEAWSYNDNDY 496
QY      525  LHDDGDNMLNMRRLSLLSLAKPLNORS--GMPLITN--SLVNRITDDNAETAAPSY 581
Db      497  NQDTNGAALAMNDGRLFSLLYTLTPINERTFGMTLIXSEYGLDTRTKNDKYGDTQPSY 556
QY      582  SFIRAHSEVQDLIRNIIRTEINPNVGVSYFTTBEIKKAFIYNKOLLATEKRYTHNTA 641
Db      557  VEVRAHSEVQTVIAQIIKEIDPTDGTFTLDQLKQAFIYNKDMNSVKNHYTHNIP 616
QY      642  LSYALLLNKSSVPRVYGDMEFTDGGYMAHTINYAEIETLLKARIKTVSGGQAKRNOQ 701
Db      617  AAYAVMLSNMESVTRVYGDFTDGGYMAHTINYAEIETLLKARIKTVSGGQAKRNOQ 676
QY      702  -----VNSEIITSVRVGKALKATO--TGDRTTTSVAVTEGNNPSL 743
Db      677  YKPSAMKAAHPDAGNVLGSEVLVSVRFGQVDSADMTGKGLAKTSQMTLISNNPEL 736
QY      744  RLKASDRVVVNNGAAHKQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIPTAAD 803
Db      737  ELDVNEEIKVNVGKHAGQAVRPLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIPTAAD 795
QY      804  IKGYPANQVSGYLVGVVPGVGAADODVRAASTAPSTD--GKSVHQAALDSRVMEFGFN 862
Db      796  IKGYPANQVSGYLVGVVPGVGAADODVRAASTAPSTD--GKSVHQAALDSRVMEFGFN 855
QY      863  FOAFATKKEEYTNVVIKAVNDKFAEMGVTDFEMAPOYVSSTDGSFLDSVIONGYAFTDRY 922
Db      856  FQDFVQKDSQYTNKKIAENTDLFKAWGVTSEFAPQVVSATDGTFLDSIIQNGYAFSDRY 915
QY      923  DLGSKPNKYGTADDLKAIALKALHSGIKVMADWVPDQYALPEKEVVTATRDVKYCTPV 982
Db      916  DLANSKNNKYSKEDLANALKAALHAGIQAIADWVPDQYIQLPEKEVVTASRDVNYGRVK 975
QY      983  AGSQIKNTLYVVDGKSSGKDDQQAQYGGAFLEELQAKYPELFARKQISTGVPMDPSPVKIKO 1042
Db      976  VDQPLVEKLYLANTKSSGKDFQAKYGGCEFLAELQKQYPEMFTAKMISTGKPIDSSVKLK 1035
QY      1043  WSACYFNGTNTILGRGAGYVLKQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGLVDF 1102
Db      1036  WSACYFNGTNTILGRGAGYVLKQATNTYFSLVSDNTFLPKSLVNPNGHGTSSSVTGLVDF 1091
QY      1103  KGYVYVYSGNQAKAFISLGNWYFDDNNGYMTGAQSLNGANYFLPSLNGIOLRAIYD 1162
Db      1092  KGIYVYVYTAGNKARSFAFTEAGNTYFYDYGHWVTGNGTINTFYFLPPLNGVLMKDAIKQ 1151
QY      1163  NGNVLSYYGNDGRYENG-----YVLF----GQ-QMRYFQON-GIMAVGLTRVHGAVQYVD 1212
Db      1152  DEKGRSVYVYGTGVNMGKGRDNEWFAMTDSKQNRFRHFDNYGFMVGLVTHGNVQYIDEE 1211
```

```
QY 1213 ASGFOAKGOFITTAGDKLRYFDRDSNQISNRFVRSKGWFLPDHNGVAVTGTVTENGQ 1272
Db 1212 ENGFOVKGEFVDQDQTRYPDQGGNVLKGF-L-NKGNWYILDDQGLVAKGAOTIKQ 1270
QY 1273 RLYF-KPNGVQAKGFIIRDANGYLRYDPSNGNEVRNRFVRNSKGWFLPDHNGIAVTGA 1331
Db 1271 KLYFDTKGVQVKGDFVTDKNGTFFYSGDTGLILGQFFSGNNAFVADENGHVAKGA 1330
QY 1332 RVNG i1336
Db 1331 KTRIG 1335

RESULT 14
ID Q8KZL5_9STRE PRELIMINARY; PRT; 1554 AA.
AC Q8KZL5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfU;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21958684; PubMed=11960691; DOI=10.1016/S0304-4165(01)00240-9;
RA Hanada N., Fukushima K., Nomura Y., Senpuku H., Hayakawa M.,
RA Mukasa H., Shiroza T., Abiko Y.;
RT "Cloning and nucleotide sequence analysis of the Streptococcus
RT sobrinus gtfU gene that produces a highly branched water-soluble
RT glucan."
RL Biochim. Biophys. Acta 1570:75-79(2002).
DR EMBL; AB089438; BAC07265.1; -; Genomic_DNA.
DR HSP; P06653; 1H8G.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 5_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ
SEQUENCE 1554 AA; 171676 MW; 6981BCC1DAE24A73 CRC64;

Query Match 44.3%; Score 3161.5; DB 2; Length 1554;
Best Local Similarity 48.5%; Pred. No. 3e-142;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;

QY 1 MEKKVRFKLKRVKKRWTVSIAASAVVLTSLSGSLVKADSTDDROQAVTESQASL----- 55
Db 1 MEKLLHYKLHKVKGHWVTIAVAS--IGLVSLVGA--GTVSABDKVANDTTAQATVGVDTG 56
QY 56 ---VTTSEAKETLTATDTSTATSOPTATVTNVSNTNSTNTANTANFVKPTTT 112
Db 57 QDQATNDANTNT--TDTTADQASNTNQDQAG--SQSNQDQAKQDPTANT-----DRNQ 108
QY 113 SQAQKTDNSDKIITTSKAVNRLTATGKFPANNNTAHPTVTDKIVPKPKIGKLKQPS 172
Db 109 ADNSQTDN-----NQATDQATS-----PATDGTISVQRDAANVATAADQEG---QTAP 153
QY 173 LSQDIIAALGNVKNIRKNGKYKYYKEDGTLOKNVALNINGTFFPD-ETGALSNTPLPS 231
Db 154 SQEKSAAAL-SLDNVKLIDGKYVYVQADGSYKKNFAITVNGQMLYFSDTSGALSTSTYS 212
QY 232 KKGNTINDNTNSQAQYNQVSTDVANFEHVHDLTAESWYRPXYLTKDGKTWTOSTEKD 291
Db 213 FSQGTIN--LVDDFSHNNKAYDSTAKSPFLVNLITANSWYRPAGILRNGQTWEASNND 270
QY 292 FRPLMTWMPDQETORQYVYNNQAQIGIHQTYNT--ATSPLOLNLAQATTQKIEEKITAE 350
Db 271 LRFVLSMWPFDKDTQVAYVYNNKYLSANETETVNTSQTVDLNKEAQS!QTKIEKITS 330
```

RESULT 15
Q55264_STRSL
ID Q55264_STRSL PRELIMINARY; PRT; 1449 AA.

```
QY 351 KNTNWLROTI SAFVYKTSANNSDSEKFPD--DHLQKGALLYSNNSKLTSAQNSYRIILNR 408
Db 331 NSTOMLRAMEAFVAAQPKWNNSTEN-FNKGDHLOGGALLYT-NSDLTPWANSRLLNLR 388
QY 409 TPTNQTGKKDPRYTADRTIGGYEFLLANDVNSNPVQAEQLNMLHFLMNFNFIYANDPD 468
Db 389 TPTQODGTK--KYPTGEGGEGYFLLSNDVNSNPVQAEQLNQLHLYLNWGDIVMGDKD 446
QY 469 ANFOSIRVDVNDVADLLQIAGDYLKAAGHGHKNDKAAANDHLSILEAWSYNDTPLYHDD 528
Db 447 ANFDGVRVDVNDVNNADLLQVYSNYFKDNYKVYTDSEANALAHISILEAWSLNDNQYNET 506
QY 529 GDNMLNMDNRLRLSLLSYLAKEPLNORSGMNPLITNSL-VNRTDDNAETAAPVSYSFIH 587
Db 507 NGTALSIDSNSRLTSLAVLTQPGQRIIDLSNISLESVKNERANDTAYGDTITPTFVRH 566
QY 588 DSEVDLIRNIIRTEINPNVVGYSPTTEEIKKAFIYNKDLIATEKKTYHYNTALSAYLL 647
Db 567 DSEVQTVIAKIVKEKIDTNSDGYTFTLQDKDAFKIYNEDMAKVNTYTHYNIIPAAYALL 626
QY 648 LTNKSSVPRVYVYGDFTDDGQYMAHKTINYBAIEFTLLKARIKYVGGQAMRNQOV-GNSE 706
Db 627 LSNMESVPRVYVYDLYTDDGQYMAKKSPPYDAIATMLQRIAYVSGGQSEEVHKVGNNG 686
QY 707 IITSRYGKGALKATDT-GDRTTRTSGVAVIEGNPNPSLRLLKASDRVNVNMGAAHKNQAYR 765
Db 687 ILSSVRYGQDLMSADDTQGTDLSTRTSGLVLTIVSNDPNLDL--GGDSLTVNMGRAHANQAYR 745
QY 766 PLLLTNDNGIKAYHSDQEAAGLVRVYNDNRGELIFTAADI KGYANPQVSGYLGWVPVGA 825
Db 746 PLLLTGDKGVQSYLKDS-D-TNIVKTDANGNLTFADDIKGYSTVDMSGYLAVWVPVGA 804
QY 826 ADQDVRVAASTAPSTDGSHVQNAALDSRVMPGFSNFQAFKATKKEEYTNVVIKNDKFP 885
Db 805 DQDQVRVAAADTNQKADGKSLKTSALDSQVIYEGFSNFQDFANNADATYTNKKAENADFF 864
QY 886 AEWGYTDFEMAPQYVSTSGSFLDSVIONGYAFTDRYDLGISKPNKYGTADDLVAKAL 945
Db 865 KKLGITSFEMAPQYVVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDLANALKAL 924
QY 946 HSKGIKVMADWVPDQYALPEKEVVTATRVKGYTPVAGSQIKNTLYVVDGKSSGKQQA 1005
Db 925 HANGIOATADWVPDQIYQLPGEEVVTAKRYSYGNPTFDAYINNALYATNTKSSGSDYQA 984
QY 1006 KYGGAFLLEOKAYPELPFARKQISTGVPMDSVKIKQWSAKYFNGTNIILRGAGYVYLKQ 1065
Db 985 QYGGAFDLDELKAKYPMDFVTNNMISTGKPIDPSTKIKQWEAKYFNGTNIILRGAGYVYLSDD 1044
QY 1066 ATNTYFSLVSDNTPELPSLVNPNHCTSSSVTGLVFDGKGYVYVYSTSGNOAKPAFISLGN 1125
Db 1045 ATGKYFTVNEGDFLPASFT---GDQNAKTFYDGTGMAYYSTSGNKAVNSFIYEGGH 1100
QY 1126 WYYPDNNGYMTGA-QSINGANYYPFLSNGIQLRNAIYDNGNKNVLSYSGNDGRRYENGYL 1184
Db 1101 YYPFDKGHWYTSYKAEDNDYELPNGIQNRDAIYQDAQNSYYGYRTGLYK----- 1155
QY 1185 FQQW-----RYFQ-NGIMAVGLTRVHGAVQYVYPDASGFAKQGFITADOK 1229
Db 1156 -GDNWYFPVNNANKTVFRYFDANNVMAIGYRNMYGQTYYPDENGFQAKGQLLTDKQ- 1213
QY 1230 LEYFDRDSNQISNRFVRSKGWFLPDHNGVAVTGTVTFFNGORLYFKP-NGVOAKGET 1288
Db 1214 THYFDRDNGAMAKNFV-NVGDDWYMDGNGNAVKGQVFNQIILYFNDETGVQVKGQFI 1272
QY 1289 RDANGLYRYDPSNGNEVRNRFVRNSKGWFLPDHNGIAVTGARVV 1334
Db 1273 TDAQRTSYDANSALKSSGFFTPNGSDWY-YAENGYYVYKQKV 1317
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 142.338 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-36
Perfect score: 7462
Sequence: 1 METKRYKMKYKGMVTVA.....EGKQVGDVAYDERLLVYR 1430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 43937871 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than, or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7462	100.0	1430	5	Aau98029 S. mutans
2	7462	100.0	1430	7	ADD93656 Streptoco
3	7462	100.0	1430	9	ADx37279 Streptoco
4	7457	99.9	1430	5	Aau98043 S. mutans
5	7456	99.9	1430	5	Aau98042 S. mutans
6	7456	99.9	1430	5	Aau98041 S. mutans
7	7451	99.9	1430	5	Aau98044 S. mutans
8	7451	99.9	1430	5	Aau98045 S. mutans
9	3869	51.8	1476	5	Aau79284 Streptoco
10	3867	51.8	1499	7	ADC54806 Protein S
11	3845.5	51.5	1475	5	Aau98038 S. mutans
12	3835.5	51.4	1475	5	Aau98037 S. mutans
13	3829.5	51.3	1475	5	Aau98036 S. mutans
14	3828.5	51.3	1475	5	Aau98035 S. mutans
15	3828.5	51.3	1475	5	Aau98034 S. mutans
16	3823.5	51.2	1475	5	Aau98033 S. mutans
17	3823.5	51.2	1475	5	Aau98032 S. mutans
18	3822.5	51.2	1475	5	Aau98031 S. mutans
19	3821.5	51.2	1475	5	Aau98040 S. mutans
20	3818.5	51.2	1475	5	Aau98030 S. mutans
21	3817.5	51.2	1475	5	Aau98027 S. mutans
22	3817.5	51.2	1475	7	ADD93654 Streptoco
23	3817.5	51.2	1475	9	ADx37277 Streptoco
24	3790.5	50.8	1475	5	Aau98039 S. mutans

25	3788.5	50.8	1577	2	AAR91047 Alpha-D-9
26	3615	48.4	1554	7	ADD93658 Streptoco
27	3615	48.4	1554	9	ADx37281 Streptoco
28	3591.5	48.1	1375	5	Aau98028 S. mutans
29	3591.5	48.1	1375	5	Aau79288 Streptoco
30	3591.5	48.1	1375	7	ADD93655 Streptoco
31	3591.5	48.1	1375	9	ADx37278 Streptoco
32	3575	47.9	1590	9	ADD93657 Streptoco
33	3575	47.9	1590	9	ADx37280 Streptoco
34	3542	47.5	1592	2	AAR32925 Glucoseylt
35	3502.5	46.9	1518	7	ADD93660 Streptoco
36	3502.5	46.9	1518	9	ADx37283 Streptoco
37	3364.5	45.1	1497	6	ABR63234 Glucansuc
38	3334.5	44.7	2835	5	ABR98574
39	3334.5	44.7	2835	6	ABR55594 Amino aci
40	3227	43.2	1477	9	ADY72696 Mutant de
41	3226	43.2	1477	9	ADY72733 Mutant de
42	3224	43.2	1477	9	ADY72732 Mutant de
43	3186.5	42.7	1527	5	Aau80055 Leuconost
44	3180.5	42.6	1527	7	ADC54807 Leuconost
45	3174	42.5	1365	7	ADD93659 Streptoco

ALIGNMENTS

RESULT 1
AAU98029
ID AAU98029 standard; protein; 1430 AA.
XX AC AAU98029;
XX DT 27-AUG-2002 (first entry)
XX DE S. mutans glucosyltransferase GTFD.
XX KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture.
XX OS Streptococcus mutans.
XX PN US2002031826-A1.
XX PD 14-MAR-2002.
XX PF 19-DEC-2000; 2000US-00740274.
XX PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
(NICH//) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX N-PSDB; ABK52940.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
as substitutes for and additions to modified starch and latexes in paper
manufacture, comprises mutations in specific positions.
XX Disclosure; Page 38-42; 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
B polypeptide having changes at position from 1448V, D457N, D567T,
K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,
I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents GTFD
XX
SQ Sequence 1430 AA;

Query Match		100.0%;	Score 7462;	DB 5;	Length 1430;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 1430;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	METKRYKMHKVKHVVAVASGLITLGTTLTGLSSVSAETPQTS	60		
DB	1	METKRYKMHKVKHVVAVASGLITLGTTLTGLSSVSAETPQTS	60		
QY	61	SSSOTDAPKTKOAEQTOAOSQANVADTSTITKETPSQNIITQANSDDKVTNTKSE	120		
DB	61	SSSOTDAPKTKOAEQTOAOSQANVADTSTITKETPSQNIITQANSDDKVTNTKSE	120		
QY	121	EAQTSEERTKQSEEAQTASSQALQAKAELTKQRTAAQENKPNVDAIENVKQIDGK	180		
DB	121	EAQTSEERTKQSEEAQTASSQALQAKAELTKQRTAAQENKPNVDAIENVKQIDGK	180		
QY	181	YYIGSDGPKKNFALTVNNKVLFPDNTGALTDTSQYQFQKGLTKLNNDYTPHNQIVNF	240		
DB	181	YYIGSDGPKKNFALTVNNKVLFPDNTGALTDTSQYQFQKGLTKLNNDYTPHNQIVNF	240		
QY	241	ENTSLETIDNYTADSWRPDKILKNGKTWTASSEDRLPLMSWPKQTOIAYLNYNN	300		
DB	241	ENTSLETIDNYTADSWRPDKILKNGKTWTASSEDRLPLMSWPKQTOIAYLNYNN	300		
QY	301	QOGLGTGENTYADSSQESLNLAQTVQVKIETKISQTOQWLRLDIINSFVKTPQNNWSQ	360		
DB	301	QOGLGTGENTYADSSQESLNLAQTVQVKIETKISQTOQWLRLDIINSFVKTPQNNWSQ	360		
QY	361	TESDTSAGEKHQLOGGALLYSNDKTA YANSYRLLNRTPTTSQTKPKYFEDNSSGGYDF	420		
DB	361	TESDTSAGEKHQLOGGALLYSNDKTA YANSYRLLNRTPTTSQTKPKYFEDNSSGGYDF	420		
QY	421	LIANDIDNSNPVQAEQLNWLHYLMNYGSIIVANDPEANPDGVRVDAVNNVADLLQIASD	480		
DB	421	LIANDIDNSNPVQAEQLNWLHYLMNYGSIIVANDPEANPDGVRVDAVNNVADLLQIASD	480		
QY	481	YLKAHYGVDKSEKNAINHLSILEAWSDDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE	540		
DB	481	YLKAHYGVDKSEKNAINHLSILEAWSDDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE	540		

Qy	541	KDASNKNFIRSGLEPVIITNSLNNRSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN	600
Db	541	KDASNKNFIRSGLEPVIITNSLNNRSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN	600
Qy	601	PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPATAYALMLSNKDSITRLYYGDMYS	660
Db	601	PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPATAYALMLSNKDSITRLYYGDMYS	660
Qy	661	DDGQYMATKSPYYDAIDTLLKARIKAYAAAGQDMKITYYVEGDKSHMDWDYTGVLTSVRYGT	720
Db	661	DDGQYMATKSPYYDAIDTLLKARIKAYAAAGQDMKITYYVEGDKSHMDWDYTGVLTSVRYGT	720
Qy	721	GANEATDQGSSEATKTQGMAVITSSNNPSLKLQNDKVIYNMGAHKNQYRPLLLTTKQGL	780
Db	721	GANEATDQGSSEATKTQGMAVITSSNNPSLKLQNDKVIYNMGAHKNQYRPLLLTTKQGL	780
Qy	781	TSYTSDDAAAKSLYRKTNDKGLVFPDASDIQGYLNPQVSGYLAVWVPVCGASNDQDVRVAAS	840
Db	781	TSYTSDDAAAKSLYRKTNDKGLVFPDASDIQGYLNPQVSGYLAVWVPVCGASNDQDVRVAAS	840
Qy	841	NKANATQGVYESSSALDSQLIYEGFSNFQDFVTXKSDYTNKKIAQNVOLFQSGWVTSFEM	900
Db	841	NKANATQGVYESSSALDSQLIYEGFSNFQDFVTXKSDYTNKKIAQNVOLFQSGWVTSFEM	900
Qy	901	APQYVSSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSGQODMINAVKALHKSIGIQVIAD	960
Db	901	APQYVSSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSGQODMINAVKALHKSIGIQVIAD	960
Qy	961	WPQDIYNLPKGVVTVATRVNDYGBYRDKSEIKNTLYAANTKSNKGDYQAKYGGAFSLSEL	1020
Db	961	WPQDIYNLPKGVVTVATRVNDYGBYRDKSEIKNTLYAANTKSNKGDYQAKYGGAFSLSEL	1020
Qy	1021	AAKYPISIFNRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG	1080
Db	1021	AAKYPISIFNRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG	1080
Qy	1081	NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDDNNGHMVYGL	1140
Db	1081	NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDDNNGHMVYGL	1140
Qy	1141	QQLNGEVQYFLSNGVOLRESFLENADGSKNYFHLGNRYSGYYSFDDNDSKWRYFPDASGV	1200
Db	1141	QQLNGEVQYFLSNGVOLRESFLENADGSKNYFHLGNRYSGYYSFDDNDSKWRYFPDASGV	1200
Qy	1201	MAVLKTTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFPANDKNGDWWYL	1260
Db	1201	MAVLKTTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFPANDKNGDWWYL	1260
Qy	1261	NSDGIALVGVQTINGKTYFPQDQKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN	1320
Db	1261	NSDGIALVGVQTINGKTYFPQDQKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN	1320
Qy	1321	WYFSGDGVAVTGSQTIAGKKLYPASDGKQVKGSGFVYNGKVHYHADSGELQVNRFEAD	1380
Db	1321	WYFSGDGVAVTGSQTIAGKKLYPASDGKQVKGSGFVYNGKVHYHADSGELQVNRFEAD	1380
Qy	1381	KDGNWYLLDSNGEALTQSGRINDQVRVFTFREGKQVKGDVAYDERLLVYR	1430
Db	1381	KDGNWYLLDSNGEALTQSGRINDQVRVFTFREGKQVKGDVAYDERLLVYR	1430
RESULT 2			
ADD93656			
ID	ADD93656 standard; protein; 1430 AA.		
XX			
AC	ADD93656;		
XX			
DT	29-JAN-2004 (first entry)		
XX			
DE	Streptococcus mutans glucosyltransferase-D.		
XX			
KW	Glucosyltransferase enzyme; vaccine; anticaries; epitope; immunogen.		

XX OS Streptococcus mutans.
XX PN WO2003075845-A2.
XX PD 18-SEP-2003.
XX PF 07-MAR-2003; 2003WO-US006962.
XX PR 07-MAR-2002; 2002US-0363209P.
XX PR 08-AUG-2002; 2002US-0402483P.
XX PA (FORS-) FORSYTH INST.
XX PI Smith DJ, Taubman MA;
XX DR WPI; 2003-845091/78.
XX PT Composition useful as vaccines for dental caries comprises a fragment of
XX PT a glucan binding protein-B binding to a major histocompatibility complex
XX PT class II protein.
XX PS Claim 16; Page 13-14; 49pp; English.
XX CC The present sequence is the protein sequence of Streptococcus mutans
XX CC glucosyltransferase-D (GTF-D). Peptide fragments of GTF-D, especially
XX CC from the catalytic domain of the polypeptide, can be used in immunogenic
XX CC compositions and subunit vaccines for dental caries. These compositions
XX CC comprise a major histocompatibility complex (MHC) class II protein-
XX CC binding peptide from S. mutans glucan binding protein-B (GbpB)
XX CC covalently linked with a peptide fragment of a streptococcal
XX CC glucosyltransferase. The compositions are used in a claimed method of
XX CC eliciting production of an antibody in a mammal. Diepitopic or
XX CC multipitopic polypeptides can be prepared synthetically or by
XX CC recombinant DNA technology. Antibodies raised against MHC class II
XX CC binding fragments of GbpB can be used in passive immunisation.
XX SQ Sequence 1430 AA;
Query Match 100.0%; Score 7462; DB 7; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 METKRYKMHKVKHWTVAASGLITLGTLLGSSVSAETEQOTSDDKVVTKQSEDDKAA 60
DB 1 METKRYKMHKVKHWTVAASGLITLGTLLGSSVSAETEQOTSDDKVVTKQSEDDKAA 60
QY 61 SESSQTDAPKTKAQTEQTAQSQANVADTSTSIKETPSQNTTQANSDDKVTNTKSE 120
DB 61 SESSQTDAPKTKAQTEQTAQSQANVADTSTSIKETPSQNTTQANSDDKVTNTKSE 120
QY 121 EAQTSSEERTKQSEEAQTASSQALTOAKALTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
DB 121 EAQTSSEERTKQSEEAQTASSQALTOAKALTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
QY 181 YYYIGSDGQPKGNFALTVNNKVLVFDKNTGALTDSQYQKGLTKLNDYTPHNOIVNF 240
DB 181 YYYIGSDGQPKGNFALTVNNKVLVFDKNTGALTDSQYQKGLTKLNDYTPHNOIVNF 240
QY 241 ENTSLTIDNVYVADSWYRPRKIDLNKGTWTWASSSDLRPLMSWPPKQTOIAYLNTMN 300
DB 241 ENTSLTIDNVYVADSWYRPRKIDLNKGTWTWASSSDLRPLMSWPPKQTOIAYLNTMN 300
QY 301 QOGLGTGENYADSSQESLNLAQTVQVKIETKISQTOQTQWLRIINSFVKTPQNNWSQ 360
DB 301 QOGLGTGENYADSSQESLNLAQTVQVKIETKISQTOQTQWLRIINSFVKTPQNNWSQ 360
QY 361 TESDTSAGEKHLQGLGALLYNSDKTAVANDSYRLNRTPTSQTKPKYFEDNSSGGYDF 420
DB 361 TESDTSAGEKHLQGLGALLYNSDKTAVANDSYRLNRTPTSQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVVOAEQLNLWHLNMYGSIIVANDPEANFDGVRVDVNDVNDLLQIASD 480

DB 421 LLANDIDNSNPVVOAEQLNLWHLNMYGSIIVANDPEANFDGVRVDVNDVNDLLQIASD 480
QY 481 YLKAHYGVDDKSEKNAIHLISILEANSNDPQYNKDTKGALPLDNKLRSLLYALTRPLE 540
DB 481 YLKAHYGVDDKSEKNAIHLISILEANSNDPQYNKDTKGALPLDNKLRSLLYALTRPLE 540
QY 541 KDASNKNEIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAAQIN 600
DB 541 KDASNKNEIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPATAYALMSKSDITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPATAYALMSKSDITRLYYGDMYS 660
QY 661 DDQYMATKSPYYDAIDTLKARIKYAAGSQDMKITVYEGDKSHMDWDVTGVLTVRYGT 720
DB 661 DDQYMATKSPYYDAIDTLKARIKYAAGSQDMKITVYEGDKSHMDWDVTGVLTVRYGT 720
QY 721 GANEATDQSEATKTQGMVITSNPNLSKLNDQKVINVMGAHKNQYRPLLLTTKDG 780
DB 721 GANEATDQSEATKTQGMVITSNPNLSKLNDQKVINVMGAHKNQYRPLLLTTKDG 780
QY 781 TSYTSDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVVPVGASDNDQVRVAAS 840
DB 781 TSYTSDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVVPVGASDNDQVRVAAS 840
QY 841 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSGVTSFEM 900
DB 841 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSGVTSFEM 900
QY 901 APQYVSSSEDSGLDSIIQNGYAFEDRYDIAMSKNNKYGSGQDMINAKALHKSIGIOVIAD 960
DB 901 APQYVSSSEDSGLDSIIQNGYAFEDRYDIAMSKNNKYGSGQDMINAKALHKSIGIOVIAD 960
QY 961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSGKDYQAKYGGAFSEL 1020
DB 961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSGKDYQAKYGGAFSEL 1020
QY 1021 AAKYPSIENRTOISNGKKIDPSEKITAMKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG 1080
DB 1021 AAKYPSIENRTOISNGKKIDPSEKITAMKAKYFNGTNIILGRGVGYVLKDNASDKYFELKG 1080
QY 1081 NQTYLPLKQMTNKEASTGFVNDGNGMTFYSTSGYQAQNSFVQDAKGNWYFDDNNGHVVYGL 1140
DB 1081 NQTYLPLKQMTNKEASTGFVNDGNGMTFYSTSGYQAQNSFVQDAKGNWYFDDNNGHVVYGL 1140
QY 1141 QOLNGEVQYFLSNGVOLRESFLENADGSKNYFGLGNRYSGYYSFDNDSKWRYFDASGV 1200
DB 1141 QOLNGEVQYFLSNGVOLRESFLENADGSKNYFGLGNRYSGYYSFDNDSKWRYFDASGV 1200
QY 1201 MAVGLKTINGTOYEDQDGYQVKGAWITGSDGKKYFDDGSGMAVNRFANDKNGDWYVL 1260
DB 1201 MAVGLKTINGTOYEDQDGYQVKGAWITGSDGKKYFDDGSGMAVNRFANDKNGDWYVL 1260
QY 1261 NSDGLALGVOTINGKTYFQDQKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320
DB 1261 NSDGLALGVOTINGKTYFQDQKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320
QY 1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKGSVFYNGKVHYHADSSELQVNRFEAD 1380
DB 1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKGSVFYNGKVHYHADSSELQVNRFEAD 1380
QY 1381 KQGNWYLLDSNGEALTGSRINDQVFPFTREGKQVKGVDVYDERLLVYR 1430
DB 1381 KQGNWYLLDSNGEALTGSRINDQVFPFTREGKQVKGVDVYDERLLVYR 1430
RESULT 3
ADX37279
ID ADX37279 standard; protein; 1430 AA.
XX
AC ADX37279;
XX

DT 21-APR-2005 (first entry)
XX Streptococcus mutant glucan binding protein B variant #8.
DE immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX Streptococcus mutans.
OS US2005031633-A1.
XX 10-FEB-2005.
XX 09-MAR-2004; 2004US-00797821.
XX 13-APR-1998; 98US-0081550P.
XX 08-JAN-1999; 99US-0115142P.
XX 12-APR-1999; 99US-00290049.
XX 07-MAR-2002; 2002US-0363209P.
XX 08-AUG-2002; 2002US-0402483P.
XX 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH J J.
PA (TAUB/) TAUBMAN M A.
XX Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
DR
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 7; SEQ ID NO 36; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
SQ Sequence 1430 AA;
Query Match 100.0%; Score 7462; DB 9; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 METKRYKHKYKHWVAVASGLITLTGTLGSSVSAETEQOTSDDKVVTKSEDDKAA 60
DB 1 METKRYKHKYKHWVAVASGLITLTGTLGSSVSAETEQOTSDDKVVTKSEDDKAA 60
QY 61 SESSQTDAPKTKQAOTEQTAQSQANVADTSTSTKETPSQNIITQANSDDKVTNTKSE 120
DB 61 SESSQTDAPKTKQAOTEQTAQSQANVADTSTSTKETPSQNIITQANSDDKVTNTKSE 120
QY 121 EAQTSERTKQSEEAQTASSQALTAQKALTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
DB 121 EAQTSERTKQSEEAQTASSQALTAQKALTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
QY 181 YYYIGSDGQPKNFALTNNKVLVDFDKNTGALTDTSTQYQKGLTKLNDYTPHNIQVNF 240
DB 181 YYYIGSDGQPKNFALTNNKVLVDFDKNTGALTDTSTQYQKGLTKLNDYTPHNIQVNF 240
QY 241 ENTSLETIDNYTADSWYRPDKILKNGKKTWTASSSDRLPLMSWMPDKQTOIAYLNYN 300
DB 241 ENTSLETIDNYTADSWYRPDKILKNGKKTWTASSSDRLPLMSWMPDKQTOIAYLNYN 300
QY 301 QOGLGTGENYTADSSQESLNLAQAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNNWSQ 360
DB 301 QOGLGTGENYTADSSQESLNLAQAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNNWSQ 360

QY 361 TESDTSAAGEKDLQGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKFYEDNSSGGYDF 420
DB 361 TESDTSAAGEKDLQGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKFYEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQLNWLHLNYGSIIVANDPEANFDGVRVDAVNVNADLLQIASD 480
DB 421 LLANDIDNSNPVQAEQLNWLHLNYGSIIVANDPEANFDGVRVDAVNVNADLLQIASD 480
QY 481 YLKAHYGVDKSEKNAINHLISLEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB 481 YLKAHYGVDKSEKNAINHLISLEAWSNDNDPQNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
QY 541 KDAASNKEIRSGLEPVIITNSLNNRSAGEKNSERMANYIFIRAHDEVOVIAKIIKAQIN 600
DB 541 KDAASNKEIRSGLEPVIITNSLNNRSAGEKNSERMANYIFIRAHDEVOVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPATAYALMLSNKDSITRLYGDMS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPATAYALMLSNKDSITRLYGDMS 660
QY 661 DDGQYMATKSPYYDAITLLKARIKYAAGGQDMKITTYEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDGQYMATKSPYYDAITLLKARIKYAAGGQDMKITTYEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQGSBATKTQGMAVITSSNPSLKLQNDKVIIVNMGAHKNQEVRLPILLTKDGL 780
DB 721 GANEATDQGSBATKTQGMAVITSSNPSLKLQNDKVIIVNMGAHKNQEVRLPILLTKDGL 780
QY 781 TSYTSDDAAKSLYRKNTDKGELVFPDASDIQGYLNPQVSGYLAVWVPVPGASDNQDVRVAAS 840
DB 781 TSYTSDDAAKSLYRKNTDKGELVFPDASDIQGYLNPQVSGYLAVWVPVPGASDNQDVRVAAS 840
QY 841 NKANATQGVYESSSALDSQLIYEGFSNFQDFVTKDSYTNKKAQNVQLFKSWGVTSPM 900
DB 841 NKANATQGVYESSSALDSQLIYEGFSNFQDFVTKDSYTNKKAQNVQLFKSWGVTSPM 900
QY 901 APOYVSSSDGFLDSIIQNGVAFEDRYDLANSKNNKYGSGQDDMINAVKALHKSGLQVAD 960
DB 901 APOYVSSSDGFLDSIIQNGVAFEDRYDLANSKNNKYGSGQDDMINAVKALHKSGLQVAD 960
QY 961 WVPDQIYNLPGKEVVVTAIRVNDYGEYRKDSEIKNTLYAANTKNSGDKYQAKYGGAFSL 1020
DB 961 WVPDQIYNLPGKEVVVTAIRVNDYGEYRKDSEIKNTLYAANTKNSGDKYQAKYGGAFSL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGVYLDKNASDKYFELKG 1080
DB 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGVYLDKNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVODAKGNWYFDNNGHMVYGL 1140
DB 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVODAKGNWYFDNNGHMVYGL 1140
QY 1141 QOLNGEVQYFLSNGVQLRESFLENADGSKNYPFGLHGNRYSGYYSFNDNSKWRYPDASGV 1200
DB 1141 QOLNGEVQYFLSNGVQLRESFLENADGSKNYPFGLHGNRYSGYYSFNDNSKWRYPDASGV 1200
QY 1201 MAVGLKTINGNTQYPDQGYQVKGAWITGSDGKKRYFDDGSGMAVNRFPANDKNGDWYVL 1260
DB 1201 MAVGLKTINGNTQYPDQGYQVKGAWITGSDGKKRYFDDGSGMAVNRFPANDKNGDWYVL 1260
QY 1261 NSDGTALVGVTINGKTYTFGDDGKQIKGKIITDNGKLKYFLANSSELARNIPATDSQNN 1320
DB 1261 NSDGTALVGVTINGKTYTFGDDGKQIKGKIITDNGKLKYFLANSSELARNIPATDSQNN 1320
QY 1321 WYFSGDGVAVTGSQTIAGKKLYFASDGKQVKGSEFVYNGKHYHYHADSSELQVNRFEAD 1380
DB 1321 WYFSGDGVAVTGSQTIAGKKLYFASDGKQVKGSEFVYNGKHYHYHADSSELQVNRFEAD 1380
QY 1381 KDGNNYLDNSNGEALTGSQRINDQVFFTREKQVKGDAVYDERLLVYR 1430
DB 1381 KDGNNYLDNSNGEALTGSQRINDQVFFTREKQVKGDAVYDERLLVYR 1430

RESULT 4

AAU98043
 ID AAU98043 standard; protein; 1430 AA.
 XX
 AC AAU98043;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFD mutant N471D.
 XX
 KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 XX
 FT Key Location/Qualifiers
 FT Misc-difference 471 /note= "Wild-type Asn substituted by Asp"
 XX
 XX US2002031826-A1.
 XX
 XX 14-MAR-2002.
 XX
 XX 19-DEC-2000; 2000US-00740274.
 XX
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX
 PI Nichols SE; }
 XX
 XX WPI; 2002-414332/44.
 XX
 XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 PS Claim 36; Page; 4app; English.
 XX
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 1448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes F1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan

CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFD mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFD sequence appearing as AAU98029 and the information in
 CC claim 36
 XX

SQ Sequence 1430 AA;

Query Match 99.9%; Score 7457; DB 5; Length 1430;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1429; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 METKRRYKMKVKKGHWTVAVASGLITLTGTTTIGSSVSAETEQTSDKVVTKSEDDKAA 60
 DB 1 METKRRYKMKVKKGHWTVAVASGLITLTGTTTIGSSVSAETEQTSDKVVTKSEDDKAA 60
 QY 61 SSSQTDAPKTKQAQTEQTAQSQANVADTSTSITKETPSQNTTQANSDDKTVTNTKSE 120
 DB 61 SSSQTDAPKTKQAQTEQTAQSQANVADTSTSITKETPSQNTTQANSDDKTVTNTKSE 120
 QY 121 EAQTSSEERTKQSEEAQTASSQAALTOAKAELTKOROTAAQENKNPVDLAAIPNVKQIDGK 180
 DB 121 EAQTSSEERTKQSEEAQTASSQAALTOAKAELTKOROTAAQENKNPVDLAAIPNVKQIDGK 180
 QY 181 YYYIGSDGQPKKFALTNNKVLTFDKNTGALTDTTSQYQFKGLTKLNNDYTPHNQIVNF 240
 DB 181 YYYIGSDGQPKKFALTNNKVLTFDKNTGALTDTTSQYQFKGLTKLNNDYTPHNQIVNF 240
 QY 241 ENTSLETIDNYVTADSWTRPKDILKNGKTTWTASSEDRLRPLMSWMPDKQTQIAYLNYN 300
 DB 241 ENTSLETIDNYVTADSWTRPKDILKNGKTTWTASSEDRLRPLMSWMPDKQTQIAYLNYN 300
 QY 301 QOGLGTGENYTDSSQESLNLAQTVOVKIETKISQTOQTOWLRDLINSFVKTOPNNSQ 360
 DB 301 QOGLGTGENYTDSSQESLNLAQTVOVKIETKISQTOQTOWLRDLINSFVKTOPNNSQ 360
 QY 361 TESDTSAGEKHLOGGALLYNSDKTAVANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
 DB 361 TESDTSAGEKHLOGGALLYNSDKTAVANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
 QY 421 LLANDIDNSNPVQAEQLNHLWHLNLYSGIVANDPEANPDGVRVDAVDNVDADLLQIASD 480
 DB 421 LLANDIDNSNPVQAEQLNHLWHLNLYSGIVANDPEANPDGVRVDAVDNVDADLLQIASD 480
 QY 481 YLKAHYGVDKSEKNAINHLSLILEANSDNDPNYKDTKGQALPIDNKLRLSLLYALTRPLE 540
 DB 481 YLKAHYGVDKSEKNAINHLSLILEANSDNDPNYKDTKGQALPIDNKLRLSLLYALTRPLE 540
 QY 541 KDASNKNIRSGLEPVIITNSLNRSABGNKSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
 DB 541 KDASNKNIRSGLEPVIITNSLNRSABGNKSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
 QY 601 PKTDGLTFTLDELKQAFKIYNEDMROAKCKYQTSNIPTAYALMLSNKDSITRLYYGDMYS 660
 DB 601 PKTDGLTFTLDELKQAFKIYNEDMROAKCKYQTSNIPTAYALMLSNKDSITRLYYGDMYS 660
 QY 661 DDGQYMATKSPYDAIDTLLKARIKYAAGQDMKITTYEGDKSHMDWDYTGVLTSVRYGT 720
 DB 661 DDGQYMATKSPYDAIDTLLKARIKYAAGQDMKITTYEGDKSHMDWDYTGVLTSVRYGT 720
 QY 721 GANEATDQGSSEATKTQGMVITSSNPSLKNQNDKVINMGAHKNQEVRLPILLTTKDL 780
 DB 721 GANEATDQGSSEATKTQGMVITSSNPSLKNQNDKVINMGAHKNQEVRLPILLTTKDL 780
 QY 781 TSYTSDAAAKSLYRKNTDKGELVDFDSDIQGLNPNQVSGYLAVWVPVGSNDQDVRVAAS 840
 DB 781 TSYTSDAAAKSLYRKNTDKGELVDFDSDIQGLNPNQVSGYLAVWVPVGSNDQDVRVAAS 840

Db 781 TSYTSDAAAKSLRYKTNKDGLVDFDASDIQGYLNPQVSGYLAVVPVVGASDNDQDVRVAAS 840
QY 841 NKANATGQVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVLFKSWGVTSPFM 900
Db 841 NKANATGQVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVLFKSWGVTSPFM 900
QY 901 APQYSSSDGSFLDSTIQNGYAFEDRYDLAMSKNNKYGSOQDMINAVKALHKSGLQVAD 960
Db 901 APQYSSSDGSFLDSTIQNGYAFEDRYDLAMSKNNKYGSOQDMINAVKALHKSGLQVAD 960
QY 961 WYPDQIYNLPKGVEVTVATRVNDYGEYRKDSSEIKNTLYAANTKSNKDYQAKYGGAFLSL 1020
Db 961 WYPDQIYNLPKGVEVTVATRVNDYGEYRKDSSEIKNTLYAANTKSNKDYQAKYGGAFLSL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTILGRGVGYLVKDNASDKYFELKG 1080
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTILGRGVGYLVKDNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVQDAKGNWYFDDNNGHMYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVQDAKGNWYFDDNNGHMYGL 1140
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFCHLGNRYSNGYYSFDDNSKWRYPDASGV 1200
Db 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFCHLGNRYSNGYYSFDDNSKWRYPDASGV 1200
QY 1201 MAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDDGSGMAVNRFPANDKNGDWWYL 1260
Db 1201 MAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDDGSGMAVNRFPANDKNGDWWYL 1260
QY 1261 NSDGTALVCVORTINGKTYYPGDDGQIKGKIITDNGKLKYFLANSSELARNIFATDSQNN 1320
Db 1261 NSDGTALVCVORTINGKTYYPGDDGQIKGKIITDNGKLKYFLANSSELARNIFATDSQNN 1320
QY 1321 WYFSGDGVAVTGSOTIAGKLYFASDQKVKGSFVTYNGKVHYHADSGELQVNRFEAD 1380
Db 1321 WYFSGDGVAVTGSOTIAGKLYFASDQKVKGSFVTYNGKVHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYLDNSGEALTGSORINDQVFFTRFGKQVKGVDVYDERLLVYR 1430
Db 1381 KQGNWYLDNSGEALTGSORINDQVFFTRFGKQVKGVDVYDERLLVYR 1430

RESULT 5
AAU98042
ID AAU98042 standard; protein; 1430 AA.

XX AC AAU98042;

XX DT 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFD mutant T589E.

DE DE Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX OS Streptococcus mutans.

OS Synthetic.

XX FH Key Location/Qualifiers

XX Misc-difference 589

FT /note= "wild-type Thr substituted by Glu"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex.
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36

XX Sequence 1430 AA;

QY Query Match 99.9%; Score 7456; DB 5; Length 1430;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 METKRRYKQHKVKKHWTVAVASGLITLTGTTTLGSSVSAETEQQTSKVVTKSDDKAA 60
Db 1 METKRRYKQHKVKKHWTVAVASGLITLTGTTTLGSSVSAETEQQTSKVVTKSDDKAA 60
QY 61 SESSOTDAPKTKQAOTEQTQASQANVADTSTSIKTPSQNITTTQANSDDKTVNTKSE 120
Db 61 SESSOTDAPKTKQAOTEQTQASQANVADTSTSIKTPSQNITTTQANSDDKTVNTKSE 120
QY 121 EAQTSSEERTKQSEEAQTASSQALYQAKAELTKQRTAAQENKNPVDLAAIPNVKQIDGK 180
Db 121 EAQTSSEERTKQSEEAQTASSQALYQAKAELTKQRTAAQENKNPVDLAAIPNVKQIDGK 180

QY 181 YYYIGSDGPKQKQNFALTVNNKVLVYDFDKNTGALTDTSTQYQFKQGLTKLNNDYTPHNQIVNF 240
DB 181 YYYIGSDGPKQKQNFALTVNNKVLVYDFDKNTGALTDTSTQYQFKQGLTKLNNDYTPHNQIVNF 240
QY 241 ENTSLLETIDNYTADSWYRPKDILKNGKWTASSEDLRLPMLSWPDKQTOAYLNYMN 300
DB 241 ENTSLLETIDNYTADSWYRPKDILKNGKWTASSEDLRLPMLSWPDKQTOAYLNYMN 300
QY 301 QOGLGTGENYTDSSQESLNLAAQTQVQKTIETKISQTQQTQWLRLDIINSFVKTPQNNWSQ 360
DB 301 QOGLGTGENYTDSSQESLNLAAQTQVQKTIETKISQTQQTQWLRLDIINSFVKTPQNNWSQ 360
QY 361 TESDTSAGEKDHLOGGALLYSNSDKTAYANDSYRLLNRPPTSOTGPKPKYFEDNSSGGYDF 420
DB 361 TESDTSAGEKDHLOGGALLYSNSDKTAYANDSYRLLNRPPTSOTGPKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQVQALNLWLHLYMNYGSIIVANDPEANFDGVRVDAVDVNNADLLQIASD 480
DB 421 LLANDIDNSNPVQVQALNLWLHLYMNYGSIIVANDPEANFDGVRVDAVDVNNADLLQIASD 480
QY 481 YLKAHYGVDPKSEKNAIHLISLEAWSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
DB 481 YLKAHYGVDPKSEKNAIHLISLEAWSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
QY 541 KQASNKQEIIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 541 KQASNKQEIIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELQAFKLYNEDMRQAKKKTQSNIPYALMLSNKDSITRLYGDMSY 660
DB 601 PKTDGLTFTLDELQAFKLYNEDMRQAKKKTQSNIPYALMLSNKDSITRLYGDMSY 660
QY 661 DDQYMATKSPYYDAIDTLKARIKYAAGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDQYMATKSPYYDAIDTLKARIKYAAGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQGEATKTQGMNAVITSNPSSLKLNQNDKVIIVNNGAAHKNQYRPLLLTTKQGL 780
DB 721 GANEATDQGEATKTQGMNAVITSNPSSLKLNQNDKVIIVNNGAAHKNQYRPLLLTTKQGL 780
QY 781 TSYTSDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVPVPGASDNQDVRVAAS 840
DB 781 TSYTSDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVPVPGASDNQDVRVAAS 840
QY 841 NKANATQVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSWGVTSPFM 900
DB 841 NKANATQVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVQLFKSWGVTSPFM 900
QY 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSSQDMINAVKALHKSIGQVIAD 960
DB 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSSQDMINAVKALHKSIGQVIAD 960
QY 961 WVPDQIYNLPKGEVTVATRVNDYGEYRKDSSEIKNTLYAANTKSNKGDYQAKYGGAPLSL 1020
DB 961 WVPDQIYNLPKGEVTVATRVNDYGEYRKDSSEIKNTLYAANTKSNKGDYQAKYGGAPLSL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVGVYVLDKNASDKYFELKG 1080
DB 1021 AAKYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNIILGRGVGVYVLDKNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFPVQDAKGNWYFFDNNGHMVTGL 1140
DB 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFPVQDAKGNWYFFDNNGHMVTGL 1140
QY 1141 QOLNGEVQYFLSNGVQLRESFLENADGSKNYFHLGNRYNSGYYSFNDNSKWRYPDASGV 1200
DB 1141 QOLNGEVQYFLSNGVQLRESFLENADGSKNYFHLGNRYNSGYYSFNDNSKWRYPDASGV 1200
QY 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNFANDKNGDWYLL 1260
DB 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNFANDKNGDWYLL 1260

QY 1261 NSDGIALVGQTINGKTYTFEGQDGKQIKGKIITDNGKLKYLFLANSSELARNIFATDSQNN 1320
DB 1261 NSDGIALVGQTINGKTYTFEGQDGKQIKGKIITDNGKLKYLFLANSSELARNIFATDSQNN 1320
QY 1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKSFTVYNGKLVHYHADSGLQVNRFEAD 1380
DB 1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKSFTVYNGKLVHYHADSGLQVNRFEAD 1380
QY 1381 KDCNWTYLLDSNGEALTGSRINDQRVFFTRREGKQVKGVDVYDERRLLVYR 1430
DB 1381 KDCNWTYLLDSNGEALTGSRINDQRVFFTRREGKQVKGVDVYDERRLLVYR 1430
RESULT 6
AAU98041
ID AAU98041 standard; protein; 1430 AA.
XX
AC AAU98041;
XX
DT 27-AUG-2002 (first entry)
XX
S. mutans glucosyltransferase GTFD mutant T589D.
XX
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 589 /note= "Wild-type Thr substituted by Asp"
XX
US2002031826-A1.
XX
14-MAR-2002.
XX
19-DEC-2000; 2000US-00740274.
XX
07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
(NICH/) NICHOLS S E.
XX
Nichols SE;
XX
WPI; 2002-414332/44.
XX
Glucosyltransferase B or D protein useful for producing a glucan useful
as substitutes for and additions to modified starch and latexes in paper
manufacture, comprises mutations in specific positions.
XX
Claim 36; Page; 44pp; English.
XX
The invention an isolated protein comprising a glucosyltransferase (GTF)
B polypeptide having changes at position from I448V, D457N, D567T,
K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
GTF D polypeptide having changes at positions from T589D, T589E, N471D,
N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
complementary polynucleotide, a ribonucleic acid sequence encoding the
GTF mutant, an expression cassette comprising the polynucleotide operably
linked to a promoter, a vector comprising the expression cassette, host
cell introduced with the vector, a transgenic plant comprising the
vector, a seed or tuber from the transgenic plant, a paper sizing and/or
coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GFP, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GFP, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GFPD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GFPD sequence appearing as AAU98029 and the information in
CC claim 36
XX
SQ Sequence 1430 AA;

Query Match 99.9%; Score 7456; DB 5; Length 1430;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 METKRRYKMHVKYKHVTVAVASGLITLGTTLTGSSVSAETEQSDTKVVTQKSEDDKAA 60
DB 1 METKRRYKMHVKYKHVTVAVASGLITLGTTLTGSSVSAETEQSDTKVVTQKSEDDKAA 60

QY 61 SESSQTDAPKTKQAOTEQTAQOSQANVADTSTSIKETPSONITTOANSDDKTVNTKSE 120
DB 61 SESSQTDAPKTKQAOTEQTAQOSQANVADTSTSIKETPSONITTOANSDDKTVNTKSE 120

QY 121 EAQTSHERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
DB 121 EAQTSHERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180

QY 181 YYYIGSDGQPKNFALTNNKVLVYFDKNTGALTDTTSQYQFKQGLTKLNNDYTPHNQIVNF 240
DB 181 YYYIGSDGQPKNFALTNNKVLVYFDKNTGALTDTTSQYQFKQGLTKLNNDYTPHNQIVNF 240

QY 241 ENTSLETTIDNYVTADSWRPKDILKNGKTTWTASSSDLRPLLMSSWPKQTOIAYLNNYN 300
DB 241 ENTSLETTIDNYVTADSWRPKDILKNGKTTWTASSSDLRPLLMSSWPKQTOIAYLNNYN 300

QY 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTMWRDLINSFVKTPQPNWSQ 360
DB 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTMWRDLINSFVKTPQPNWSQ 360

QY 361 TESDTSAGEKHLQGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
DB 361 TESDTSAGEKHLQGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420

QY 421 LLANDIDNSNPVQAEQNLWHLVLMNYGSIIVANDPEANPDGVRVDAVNNVNDLLOIASD 480
DB 421 LLANDIDNSNPVQAEQNLWHLVLMNYGSIIVANDPEANPDGVRVDAVNNVNDLLOIASD 480

QY 481 YLKAHYGVDDKSEKNAINHLISLEANSNDPQNKDTKGAQLPDKLRLSLVYALTRPLE 540
DB 481 YLKAHYGVDDKSEKNAINHLISLEANSNDPQNKDTKGAQLPDKLRLSLVYALTRPLE 540

QY 541 KDAASKNEIRSGLEPVITNSLNNRSGKNSRMANIYIFIRAHDSQVQTVIAKIIKAQIN 600
DB 541 KDAASKNEIRSGLEPVITNSLNNRSGKNSRMANIYIFIRAHDSQVQTVIAKIIKAQIN 600

QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPATAYALMLSNKDSITRLYYGDMYS 660

DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
QY 661 DDGQYMATKSPYDAIDTLLKARIKYAAGGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDGQYMATKSPYDAIDTLLKARIKYAAGGQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQGSSEATKTQGMVITSNPDLKLNQNDKVIYVMGAAHKNQYRPLLLTTKQGL 780
DB 721 GANEATDQGSSEATKTQGMVITSNPDLKLNQNDKVIYVMGAAHKNQYRPLLLTTKQGL 780
QY 781 TSYSDAAAKSLYKRTNKGELVDPASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
DB 781 TSYSDAAAKSLYKRTNKGELVDPASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
QY 841 NKANATGVYESSSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVLFKSWGVTSPFM 900
DB 841 NKANATGVYESSSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVLFKSWGVTSPFM 900
QY 901 APQYVSSSEGGSFLLDSIIQNGYAFEDRYDLAMSKNNKYGSGQDMINAVKALHKSIGQVIAD 960
DB 901 APQYVSSSEGGSFLLDSIIQNGYAFEDRYDLAMSKNNKYGSGQDMINAVKALHKSIGQVIAD 960
QY 961 WVPDQIYNLPGKEVVVTATRVNDYGEYRKDSEIKNTLYAANTKSGKDYQAKYGGAFSLSEL 1020
DB 961 WVPDQIYNLPGKEVVVTATRVNDYGEYRKDSEIKNTLYAANTKSGKDYQAKYGGAFSLSEL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVYLVKDNASDKYFELKG 1080
DB 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVYLVKDNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGQYAKNSFVQDAKGNWYTFDNNGHMVYGL 1140
DB 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGQYAKNSFVQDAKGNWYTFDNNGHMVYGL 1140
QY 1141 QOLNCEVOYFVLSNGVQLRESFLENADGSKNYFGLHGNRYSGYYSFNDNSKWRYFDASGV 1200
DB 1141 QOLNCEVOYFVLSNGVQLRESFLENADGSKNYFGLHGNRYSGYYSFNDNSKWRYFDASGV 1200
QY 1201 MAVGLKTINGNTQYFPDQGYQVKGAWITGSDGKKRYFDDGSGNMAVNFANDKNGDWWYL 1260
DB 1201 MAVGLKTINGNTQYFPDQGYQVKGAWITGSDGKKRYFDDGSGNMAVNFANDKNGDWWYL 1260
QY 1261 NSDGIALVGVQTINGKTYTFGQDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
DB 1261 NSDGIALVGVQTINGKTYTFGQDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WYTFGSDGVAVTGSOTIAGKKLYFASDQGVKVGKSFVYNGKVHYHADSGELQVNRFEAD 1380
DB 1321 WYTFGSDGVAVTGSOTIAGKKLYFASDQGVKVGKSFVYNGKVHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYILDSNGEALTGSRINDQRFVFFTRREGQVKGVDVAYDERLLVYR 1430
DB 1381 KQGNWYILDSNGEALTGSRINDQRFVFFTRREGQVKGVDVAYDERLLVYR 1430

RESULT 7
AAU98044
ID AAU98044 standard; protein; 1430 AA.
XX AC AAU98044;
XX DT 27-AUG-2002 (first entry)
XX DE S. mutans glucosyltransferase GTFD mutant N471D/T589D.
XX KW Glucosyltransferase; GFPD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
XX OS Streptococcus mutans.
XX OS Synthetic.

FH Key | Location/Qualifiers
 FT Misc-difference 471 /note= "Wild-type Aen substituted by Asp"
 FT Misc-difference 589 /note= "Wild-type Thr substituted by Asp"
 FT
 FT US2002031826-A1.
 PN 14-MAR-2002.
 XX 19-DEC-2000; 2000US-00740274.
 XX 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX
 XX Nichols SE;
 XX
 XX WPI; 2002-414332/44.
 DR
 XX
 PT Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 XX Claim 36; Page; 44pp; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart Gloss to the paper during coating step. The present
 CC sequence represents a GTFD mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFD sequence appearing as AAU98029 and the information in
 CC claim 36
 XX
 XX Sequence 1430 AA;

Query Match	99.9%	Score 7451;	DB 5;	Length 1430;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 1428;	Conservative 1;	Mismatches 1;	Indels 0;	Gaps 0;
QY 1	METKRRYKHKVKKHVVTVAVASGLITLGTITLGGSSVSAETEOOTSQVTKSDDKAA	60		
DB 1	METKRRYKHKVKKHVVTVAVASGLITLGTITLGGSSVSAETEOOTSQVTKSDDKAA	60		
QY 61	SESSQTDAPKTKOAOQTEOTQAOQANVADTSTSTTKETPSQNTTTOANSDDKTVNTKSE	120		
DB 61	SESSQTDAPKTKOAOQTEOTQAOQANVADTSTSTTKETPSQNTTTOANSDDKTVNTKSE	120		
QY 121	EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQIDGK	180		
DB 121	EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQIDGK	180		
QY 181	YYIGSDGQPKKFPALTAVNKVLYFDKNTGALTDTTSQYQFKQGLTKLNNDYTPHNQIVNF	240		
DB 181	YYIGSDGQPKKFPALTAVNKVLYFDKNTGALTDTTSQYQFKQGLTKLNNDYTPHNQIVNF	240		
QY 241	ENTSLETIDNYVTADSWYRPKDILKNGKWTWASSSDLRPLLMWWPDKOTQIAYLNYMN	300		
DB 241	ENTSLETIDNYVTADSWYRPKDILKNGKWTWASSSDLRPLLMWWPDKOTQIAYLNYMN	300		
QY 301	QQGLGTGENYTADSSQESLNLAQTVQVKIETKISQTOQTQWLRLDIINSFVKTPQWNNSQ	360		
DB 301	QQGLGTGENYTADSSQESLNLAQTVQVKIETKISQTOQTQWLRLDIINSFVKTPQWNNSQ	360		
QY 361	TESDTSAGEKDHLOGGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKPKFEDNSSGGYDF	420		
DB 361	TESDTSAGEKDHLOGGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKPKFEDNSSGGYDF	420		
QY 421	LLANDIDNSNPVVQAEQLNWLHYLMNYSIVANDPEANFDGVRVADVADNVNADLLQIASD	480		
DB 421	LLANDIDNSNPVVQAEQLNWLHYLMNYSIVANDPEANFDGVRVADVADNVNADLLQIASD	480		
QY 481	YLKAHYGVDKSEKNAINHLSILEASDNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE	540		
DB 481	YLKAHYGVDKSEKNAINHLSILEASDNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE	540		
QY 541	KDASKNEIRSGLEPVITNSLNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN	600		
DB 541	KDASKNEIRSGLEPVITNSLNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN	600		
QY 601	PXTDGLTFTLDELKQAFKLYNEDMQAKKYYTQSNIPATAYALMNLNKDSITRLYYGDMYS	660		
DB 601	PXTDGLTFTLDELKQAFKLYNEDMQAKKYYTQSNIPATAYALMNLNKDSITRLYYGDMYS	660		
QY 661	DDGQYMATKSPYYDAIDTLLKARIKYAAGGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT	720		
DB 661	DDGQYMATKSPYYDAIDTLLKARIKYAAGGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT	720		
QY 721	GANEATDQGEATKTCGMNAVITSNPNLSKLNQNDKVIYVMGAAHKNQEVRLPILLTKDGL	780		
DB 721	GANEATDQGEATKTCGMNAVITSNPNLSKLNQNDKVIYVMGAAHKNQEVRLPILLTKDGL	780		
QY 781	TSYTSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLVAVVPVGSADNQDVRVAAS	840		
DB 781	TSYTSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLVAVVPVGSADNQDVRVAAS	840		
QY 841	NKANATGVYESSSALDSOLIYEGFSNFODFTVKDSYTNKIIAQNQLPKSGWTSFEM	900		
DB 841	NKANATGVYESSSALDSOLIYEGFSNFODFTVKDSYTNKIIAQNQLPKSGWTSFEM	900		
QY 901	APQYVSSDGGSLDSIIQNGYAFEDRYDLAMSKNNKYSQQDMINAVKALHKSIGQIVAD	960		
DB 901	APQYVSSDGGSLDSIIQNGYAFEDRYDLAMSKNNKYSQQDMINAVKALHKSIGQIVAD	960		
QY 961	WVPDQIYNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFLSL	1020		
DB 961	WVPDQIYNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFLSL	1020		
QY 1021	AAKPSIFNRQTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVYVYKDNASDKYFELKG	1080		

Db 1021 AAKYPSIFNRTQISNGKIDPSEKITAWEKAKIFNGNTILGRGVGVYLDKNASDKYFELXG 1080
QY 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFYFDNNGHMVYGL 1140
Db 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFYFDNNGHMVYGL 1140
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHIGNRYSNGYYSFDNDSKRYFDASGV 1200
Db 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHIGNRYSNGYYSFDNDSKRYFDASGV 1200
QY 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDCKRYFDDGSGNMAVNRFPANDKNGDWYVL 1260
Db 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDCKRYFDDGSGNMAVNRFPANDKNGDWYVL 1260
QY 1261 NSDGIALGVQTINGKTYFYFGDGKQIKGIITDNGKLYFLANSSELARNIFATDSQNN 1320
Db 1261 NSDGIALGVQTINGKTYFYFGDGKQIKGIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WTYFGSDGAVVNGSQTIAKGLYFASDGKQVKGSGFTYNGKHYHYHADSGELQVNRFEAD 1380
Db 1321 WTYFGSDGAVVNGSQTIAKGLYFASDGKQVKGSGFTYNGKHYHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYLDNSGEALTGSORINDQVFFFTREGKQVKGDVAYDERLLVVR 1430
Db 1381 KQGNWYLDNSGEALTGSORINDQVFFFTREGKQVKGDVAYDERLLVVR 1430

RESULT 8

AAU98045
ID AAU98045 standard; protein; 1430 AA.
AC AAU98045;
XX
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFD mutant N471D/T589E.
XX
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 471
FT /note= "Wild-type Asn substituted by Asp"
FT Misc-difference 589
FT /note= "Wild-type Thr substituted by Glu"
XX
XX US2002031826-A1.
XX
XX 14-MAR-2002.
XX
XX 19-DEC-2000; 2000US-00740274.
XX
XX 07-JUN-1995; 95US-00478704.
XX 07-JUN-1995; 95US-00482711.
XX 07-JUN-1995; 95US-00485243.
XX 16-JAN-1998; 98US-00007999.
XX 16-JAN-1998; 98US-00008172.
XX 20-JAN-1998; 98US-00009620.
XX 11-DEC-1998; 98US-00210361.
XX
XX (NICH/) NICHOLS S E.
XX
XX Nichols SE;
XX
XX WPI; 2002-414332/44.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.
XX Claim 36; Page; 44pp; English.
XX
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/D1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36
XX
SQ Sequence 1430 AA;
Query Match 99.9%; Score 7451; DB 5; Length 1430;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1428; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 METKRYKMKHKVKKHVVTVAVASGLITLGTTLGSSVSAETEQOTSDDKVVTKSEDDKAA 60
Db 1 METKRYKMKHKVKKHVVTVAVASGLITLGTTLGSSVSAETEQOTSDDKVVTKSEDDKAA 60
QY 61 SESSQTDAPKTKQAQTEQTAQSQANVADTSTSIKTEPSONITTAQNSDDKTNTKSE 120
Db 61 SESSQTDAPKTKQAQTEQTAQSQANVADTSTSIKTEPSONITTAQNSDDKTNTKSE 120
QY 121 EAQTSSEERTKQSEEAQTTASSQALTKQKQTAQKAEITKQRTAAQENKNPVDLAAIPNVKIDGK 180
Db 121 EAQTSSEERTKQSEEAQTTASSQALTKQKQTAQKAEITKQRTAAQENKNPVDLAAIPNVKIDGK 180
QY 181 YYYIGSDGPKKGFALTVNNKVLFPDKNTGALTDTFSQYQKGLTKLNDYTPHNOIYNF 240
Db 181 YYYIGSDGPKKGFALTVNNKVLFPDKNTGALTDTFSQYQKGLTKLNDYTPHNOIYNF 240
QY 241 ENTSLETIDNYVTADSWYRPKDILKNGKTTWASSSESDLRPLIAMSWWPKQTOIAYLYNN 300
Db 241 ENTSLETIDNYVTADSWYRPKDILKNGKTTWASSSESDLRPLIAMSWWPKQTOIAYLYNN 300
QY 301 QOGLGTGENYTADSSQESLNLAQTVQVKIEFKISQTOQOTQWLRLDIINSFVKTPQNNWQ 360
Db 301 QOGLGTGENYTADSSQESLNLAQTVQVKIEFKISQTOQOTQWLRLDIINSFVKTPQNNWQ 360

QY 361 TESDTAGEKXHLQGGALLYSNDSOKTAYANSYRLNRTPTSGTKPKYFEDNSSGGYDF 420
DB 361 TESDTAGEKXHLQGGALLYSNDSOKTAYANSYRLNRTPTSGTKPKYFEDNSSGGYDF 420
QY 421 LLANDINSNPVQAEOLNKLHLYLMNYGSI VANDPEANFDGVRVDVANDVNDLLOIATSD 480
DB 421 LLANDINSNPVQAEOLNKLHLYLMNYGSI VANDPEANFDGVRVDVANDVNDLLOIATSD 480
QY 481 YLKAHYGVDSKSEKNAIHLISLEAWSNDPQYNKDTKGALPIDNKLRLSLLYALTRPLE 540
DB 481 YLKAHYGVDSKSEKNAIHLISLEAWSNDPQYNKDTKGALPIDNKLRLSLLYALTRPLE 540
QY 541 KDSNKNRISGLPEVITNSLNRSAGCKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 541 KDSNKNRISGLPEVITNSLNRSAGCKNSRMANYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
QY 661 DDQYNMATSPPYYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDQYNMATSPPYYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQSEATKTQGMVITSNPDLKNDKVIYVNGAAHKNQYRPLLLTTKDEL 780
DB 721 GANEATDQSEATKTQGMVITSNPDLKNDKVIYVNGAAHKNQYRPLLLTTKDEL 780
QY 781 TSYTSDAARSLYRKNTDNGELVFDASDIQGYLNPQVSGYLAVWVPVGSADNDQVRVAAS 840
DB 781 TSYTSDAARSLYRKNTDNGELVFDASDIQGYLNPQVSGYLAVWVPVGSADNDQVRVAAS 840
QY 841 NKANATQGVYESSALDSQLIYEGFSNFODFVTKDSYTNKIKIAQNVQLFKSNGVTSFPM 900
DB 841 NKANATQGVYESSALDSQLIYEGFSNFODFVTKDSYTNKIKIAQNVQLFKSNGVTSFPM 900
QY 901 APQYVSSBGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVKALHKSIGIQUIAD 960
DB 901 APQYVSSBGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVKALHKSIGIQUIAD 960
QY 961 WYPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTYSGNDKYQAKYGGAPLSL 1020
DB 961 WYPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTYSGNDKYQAKYGGAPLSL 1020
QY 1021 AAKYPSIFNRTQISNGKIDPSEKITAWKAKYFNGTNIILGRGVYVLKDNASDKYFELKG 1080
DB 1021 AAKYPSIFNRTQISNGKIDPSEKITAWKAKYFNGTNIILGRGVYVLKDNASDKYFELKG 1080
QY 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFFDNNGHMYYGL 1140
DB 1081 NQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFFDNNGHMYYGL 1140
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSGNYGYPFNDNSKRYFDASGV 1200
DB 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSGNYGYPFNDNSKRYFDASGV 1200
QY 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKRYFDDGSGNMAVNFANDKNGDWWYL 1260
DB 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKRYFDDGSGNMAVNFANDKNGDWWYL 1260
QY 1261 NSDGIALGVQTINGKTYFFGDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
DB 1261 NSDGIALGVQTINGKTYFFGDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WTYFGSDGVAVTGTSOTIAGKKLYFASDGHQVKGSGFVYTKNGKHYHYHAGSELQVNRFEAD 1380
DB 1321 WTYFGSDGVAVTGTSOTIAGKKLYFASDGHQVKGSGFVYTKNGKHYHYHAGSELQVNRFEAD 1380
QY 1381 KDGNYWYLDNSGEALTGSRINDORVFFTRREGQVKGVDVAYDERLLAYR 1430
DB 1381 KDGNYWYLDNSGEALTGSRINDORVFFTRREGQVKGVDVAYDERLLAYR 1430

RESULT 9

AAU79284

ID AAU79284 standard; protein; 1476 AA.

XX AC AAU79284;

XX XX 13-AUG-2002 (first entry)

XX DT

XX ST Streptococcus mutans monoclonal antibody-related protein #1.

XX DE

XX KW Antibody; dental caries; water insoluble glucan synthetase; anti-caries;

XX KW glucosyl transferase-B; immunotherapy.

XX OS Streptococcus mutans.

XX XX JP20021114709-A.

XX PD 16-APR-2002.

XX PF 04-OCT-2000; 2000JP-00304889.

XX PR 04-OCT-2000; 2000JP-00304889.

XX XX (UYNI-) UNIV NIPPON.

XX DR WPI; 2002-448885/48.

XX XX

XX PT Anti-caries agent composed of a monoclonal antibody against an inhibitory

XX PT enzyme against water insoluble glucan synthetase of glucosyl transferase-

XX PT B (GTF-B) of Streptococcus mutans.

XX XX

XX PS Claim 3; Page 13-16; 28pp; Japanese.

XX XX

XX CC The invention relates to a monoclonal antibody against dental caries and

XX CC an anti-caries agent composed of a monoclonal antibody produced by

XX CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)

XX CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having

XX CC inhibitive activity against water insoluble glucan synthetase of glucosyl

XX CC transferase-B. The monoclonal antibody specifically inhibits water

XX CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl

XX CC transferase-B and is used in the immunotherapy of dental caries. This

XX CC sequence represents a Streptococcus mutans monoclonal antibody-related

XX CC protein

XX SQ Sequence 1476 AA;

Query Match

Best Local Similarity 51.8%; Score 3869; DB 5; Length 1476;

Matches 767; Conservative 229; Mismatches 394; Indels 48; Gaps 20;

QY 1 METKERYKMHKVKHMTVAVASGLITLTTGLSSVSAETEQOTSDKVVTQKSEDDKAA 60

DB 1 MDKRYKRLKRYKRWTVSVASVMTL-TTLSSGLVXADSNESKS-----QISNDSNTS 54

QY 61 SSSQTDAPKTKQAOTEQTOAQSOAN-VADTSTSTITKETPSONITTOANSDDKTVNTKS 119

DB 55 VTANESNVITEATSKQEAASSQTHVTVSSSTSVVNPKEVSNPVTGCTASNGEK 114

QY 120 EBAQTSERTKQSEEAQTASSQALTOAKAELTKORQTAAGNKNPVDLAAIPNVKQIDG 179

DB 115 LQNTTTV-DKTSEAAANNISKQT-TEADTDVIDDSNAA-----NLQILEKLPNVKEIDG 167

QY 180 KYIYIGSDGQPKFNALTVNNKLYFDKNTGALTDTTS-QYQFKQGLTKLNNDTYTHNQIV 238

DB 168 KYIYDNNKGRVNTFTLIADGKILHFE-TGAYTDTSTDTVKNKDIIVTTRSNLYKKYNQY 226

QY 239 NPENTSLTIDNYTADSWYRPKDILKNGKVTASSESDRLPRLMSWPDKOTQIAYLNY 298

DB 227 DRSAQSPFHVHLYTAESWYRPKYLLKQKWTQSTKDFRPLLMTWPDQSTQIQYNY 286

QY 299 MNQOGLGTGENYTADSSQESLNAAQTVQVKIETKISQTOQTQWLRIINSFVKTPQNN 358

DB 287 MNAQ-LGINKTYDDTSNQLNIAAATQAKIEAKITTLKNTDLRLQITSAFVKTSAWN 345

QY 359 SOTESDTSAGEKHLOGGALLYSNDK-TAYANSYRLNRTPTQTK--PKYPEDRSS 415
Db 346 SDESEKPF-----DHLQNGAVLYDNEGKLPYANSYRLNRTPTQTKKDPRTADNTI 401
QY 416 GGYDFLLANDIONSPPVQAEOLNWLHYLMNYSIVANDPEANFDGVRVDAVDNVDNADLL 475
Db 402 GGYEFLLANDVNSNPVQAEOLNWLHFLMNFNIYANDPDANFDSIRVDAVDNVDADLL 461
QY 476 QIASDYLKARYGVDSKNAIHLISLEAWSNDNDPOYNKDTKGAQLPIDNKLRLLSLLYAL 535
Db 462 QIAGDYLLKAAKGTHKNDKAANDHLSILEAWSNDPTYLHDDGDNINMDNKLRLSLLFSL 521
QY 536 TPLEKDAANKNEIRSGLEPVTINSLNRSAGKISERMANIYIFRAHDSVQTVIAKII 595
Db 522 AKPLNQ-----RSGMPLITNSLVNRTDDNAETAAPVSYSFRAHDSVQDLIRDI 573
QY 596 KQAINPKTDGLFTLDELKQAFKYNEDMRQAKKYTQSNIPAYALMLSNKDSITRLY 655
Db 574 KAEINPVVGYSTWEEIKAFIYNKOLLATEKKITHYNTALSTALLTNKSSVPRVY 633
QY 656 GMYSDGQYMATKSPYDAIDTLKARIKYAAGGQDMKITYVEGDKSHMDWDYTGVLTS 715
Db 634 GDMFTDDGOYMAKNTYEAETLLKARIKYVSGGQAMENQVGNSE-----IITS 684
QY 716 VRYGTGANRATOGSEATYTOGMVITSNPSSLNNDKVINMGAAHQEQEYRPLLT 775
Db 685 VRYGKALKATDGTDRTRTSGVAVIEGNNPSSLKASDRVVVNMGAHKNQAYRPLLT 744
QY 776 TKDGLTSYSDAAKSLYKTKNDKGBELVFDASDIQGYLNPQVSGYLAVVVPVPGASNDQV 835
Db 745 TUNGKAYHSDQEAAGLVRYTRDNGELIFTAIDIKGANPQVSGYLGWVPVGAADQV 804
QY 836 RVAASNKANATQVYBESSALDSQLIYEGFSNFDQFVTKDSYTNKKIAQNVQLFKSGV 895
Db 805 RVAASTAPSTDGKSVHQAALDSRVMEGFSNFQAFATKKEEYTNVVIKNDVKFAEMGV 864
QY 896 TSFEMAPQVSSDGSFLDSIITONGYAFEDRVDLAMSNNKYGSGQODMNAVKALHSGI 955
Db 865 TDFEMAPQVSSDGSFLDSVQNGYAFTRDYLGISKENKGTADDLVKAIALHSGI 924
QY 956 QVIADVPQIYNLPKGVVTRVNDYGEYKDSIKNTLYAANTKSGDKYQAKYGA 1015
Db 925 KVMADVPQMTAFPEKEVVTATRVKTYGTPVAGSQIKMLTVDGKSGKQQAQYGA 984
QY 1016 FLSELAQYPSIFNRTOISNGKKIDPSEKITAFAKYPNGTILGRGVGYVLKONASDKY 1075
Db 985 FLEELQAKYPELFARKQISTGVPMDDPSVKIKQWSAKYFNGTNILGRGAGYVLKQATNTY 1044
QY 1076 FELKGNQ--TYLPKQMTNKEASTGFVNDGNGMTFTSTGYOAKNSFVQDQAGNWFYDNN 1133
Db 1045 FNISDNKEINFLPKTLNLDQSDQVGSYDGKGYVYTSSTGYOAKNFTISEG-DKWFYFDNN 1103
QY 1134 GHWYGLQQLNGEVOYFLSNGVQLRESFLENADGSKNYFGLGNRYNGYISFNDNSKWR 1193
Db 1104 GHWVTAQASINGVNYFLSNGVQLRLDAILKNEEDGTAYVYNGDORRYENGYOF-MSGWR 1162
QY 1194 YFDASGVMAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKGRYFDDGSGNMAVNRFANDK 1253
Db 1163 HFN-NGEMSGLTVIDGVOYFDEMGYQAKGFVTTADGKIRYFDQSGNMYNRFIENE 1221
QY 1254 NGWYILNSDGLALGVQTINGKTYFFQDQKQIKGIITD-NGKKUYFLANSBELARNI 1312
Db 1222 EGKWLVLGDEGAVTGSOTINGQHLVFRANGVQVKGFEVTDHHRISYDDGNSGQIRNR 1281
QY 1313 FATDSNNWYFSGDCAVGTGQTIAGKKLYFASDGKQVKGSPVTYN-GKVHYHYADSGE 1371
Db 1282 FVRNAQGMFYFDNNGYAVTGARTINGQHLVFRANGVQVKGFEVTDRIYGRISYDNGSGD 1341
QY 1372 LQVNRPEADKGNWYILNSDGEALITGSRINQVRVFTREGQVKGVDVAYDERLLVY 1429
Db 1342 QIRNRFPVRNAQGMFYFDNNGYAVTGARTINGQHLVFRANGVQVKGFEVTDRIYGRISY 1399

RESULT 10
ADC54806
ID ADC54806 standard; protein; 1499 AA.
XX AC ADC54806;
XX AC ADC54806;
DT 18-DEC-2003 (first entry)
XX Protein Seq ID11 related to L mesenteroides dextran sucrose protein.
XX dextran sucrose; active centre zone; glucan; polyaccharide; dextran;
KW D-glucose; starch; cellulose; glucan manufacture; transduction;
KW enzyme-reaction product.
XX Unidentified.
XX JP2003111590-A.
XX 15-APR-2003.
XX 03-OCT-2001; 2001JP-00307067.
XX 03-OCT-2001; 2001JP-00307067.
XX (DOKU-) DOKURITSU GYOSEI HOJIN SHOKUJIN SOGO KEN.
XX WPI; 2003-735670/70.
XX P-PSDB; ADC54814.
XX Novel modified dextran sucrose which exchanges one site of active center
zone of dextran sucrose for active center zone of different types of
dextran sucrose, useful for manufacturing glucan.
XX Example 1; SEQ ID NO 11; 28pp; Japanese.
XX This invention relates to a modified dextran sucrose (DS) exchanging one
site of the active centre zone of a dextran sucrose for the active centre
zone of different types of dextran sucrose. Glucan is a polysaccharide
(for example dextran) which uses D-glucoses, such as a starch and a
cellulose, as a structural unit. The modified enzyme of the invention is
useful in the manufacture of glucan. The selection of the active centre
area of DS which carries out transduction, enables changes in structure
and character of an enzyme-reaction product and their application to
various uses. The present sequence is that of a protein which is related
to the leuconostoc mesenteroides dextran sucrose protein and which was
used during the exemplification of the invention.
SQ Sequence 1499 AA;
Query Match 51.8%; Score 3867; DB 7; Length 1499;
Best Local Similarity 53.4%; Pred. No. 7e-192;
Matches 775; Conservative 202; Mismatches 415; Indels 58; Gaps 19;
QY 17 VTVAVASGLITLITGLSSV-----SARTEQOTSDKVVY--QKSEDDKAA--SESSQTD 67
Db 36 VVAVAVQSNLDATSDKSIITDDKAAATATSTDDKATTTADTTDDKAAATATSTDD 95
QY 68 APKTKQAQTEQTAQOANVADTSTSTKETPSQN--ITTOANS-DDKVTYN--TKSEEA 122
Db 96 KATTTAATSTDDKATTAATSTDDKAAATATSTDDKATTTAATSTDDKATTTAATSTDDK 155
QY 123 QTSERTQSEEAQTA-----SSOALTQAKAELTKQRTAAQENK 163
Db 156 ATTTAATSTDDKAAATATSTDDKATTTAATSTDDKATTTAATSTDDKATTTAATSTDDK 215
QY 164 NPVDLAAI-----PNVKQIDGKYIYIGSDQPKKNFALTVNNKLVFDKNTGALTSTOYQ 219
Db 216 TADTTAALEASSNKLTIDGKTYIYDDDDQVKNFATVIDGKLVIFDKETGALLATNDYQ 275
QY 220 FKQGLTKLNDYTPHNQIVNFENTSLTIDNYVTADSWYRPKDILKNGKWTWASSEDLR 279
Db 276 FLEGITSENNTYTEHNASVGTTSYTNVDGYLTADSWYRPKDILVNGQWESSKDDDLR 335

QY 239 NFENTSLTIDNYVTADSWRPKIDILKNGKTTWTASSESDIRPLMLSWPDKQTOIAYLNY 298
Db 227 DRSAQSPGHDVHYLTAEBSWYRPKYLKDKGTWTSTKDFRPLMLTWPQDETQRYVNY 286
QY 299 MNOQGLGTGNYTADSSQESLNLAQOTVQVKIEKISOTQOTWLRI INSFVKTOPNWN 358
Db 287 MNAQ-LGINKTYDDTSNQLQINAAATIQAKIEAKITTLKNTDWLROTISAFVKTSQAWN 345
QY 359 SOTESDTSAGEKHLOGGALLYNSDK-TAYANSYRLNRTPTSTQTK--PKYFEDNSS 415
Db 346 SDSEKPPD---DHLQNGAVLYDNEGKLTYPYANSYRLNRTPTNQTGKKDPRTADNTI 401
QY 416 GGYDFILLANDINSNPVQAEOLNWLHYLANYGSIIVANDPEANPDGVVRDAVNANADLL 475
Db 402 GGYEFFLLANDVNSNPVQAEOLNWLHFLNFGNIYANDPDANFDSIRVDVNDVADLL 461
QY 476 QIASDYLKAHYGVDKSEKNAIHLISLEAUSDNDPQYKDKTGAQLPIDNKLRLSLLYAL 535
Db 462 QIAGDYLYKAAKGTHKNDKAAANDHLSLEAUSDNDPTYLHDDGNMINDNKLRLSLFL 521
QY 536 TRPLEKXDAKNKIBIRSGLEPVIITNSLNRSAEGKNSERMANYIFIRAHDSVQTVIAKII 595
Db 522 AKPLNQ-----RSGMNPILINSLVNRDDNAETAAPVPSYFIRAHDSVQTVIAKII 573
QY 596 KAOINPKTDGLTFLDELKQAFKIYNEDMRQAKKYYTQSNIPYAYAMLNKNDSITRLYY 655
Db 574 KAEINPNVGVSTBEIKAFIYNNKDLATEKTYHTYNTALSYALLTNKSVPRVY 633
QY 656 GDMYSDGQVMAKSPYDAIDTLKARIKAYAGGQDKITYVEGDKSHMDWDTYGVILTS 715
Db 634 GDMFTDGGVMAHKTINYEATIELLKARIKYVSGQAMRNQQVGNSE-----IITS 694
QY 716 VRYGTGANEATDQGSSEATKTQGMVITSNPISLKLQNQNDKVIYVMGAHKNQYRPLLLT 775
Db 685 VRYGKALKATDGTDRTRTSGVAVIEGNNPSRLKASDRVNVNMGAAHKNQYRPLLLT 744
QY 776 TKDGLTSYSDAAKSLYKNTDKGELVFPDASDIQGYLNPQVSGYLAVVVPVQASNDQV 835
Db 745 TONGIKAYHSDQEAAGLVRYTNDRGELIETAAIDIKGYANPQVSGYLGVVVPVGA--LI 801
QY 836 RVAASNAKANATQV--YESSALDSQLIYEGESNFODFVTKSDYNTKTKIAQNVQPKSW 893
Db 802 KMFALRLAPHOQMAVSHQNAALDSRVMEGFSNFOAFATKBEETNVVIKRVDPFAEW 861
QY 894 GVTSEFAPQYVSSSDGSLDSIIONGYAFEDRYDLAMSKNNKYGSQOQMINAVKALHS 953
Db 862 GVTDFEAPQYVSSDGSFLDSVIONGYAFETDRYDLGSKPNKYGTADDLVKAIALHSK 921
QY 954 GIOVDWVPDQIYNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAAANTKSNKQYQAKYG 1013
Db 922 GIKWMADWVPDQWYAFPEKEVVTATRVNDYKGTVPVAGSQIKNTLYVVDGKSSGKQQAKYG 981
QY 1014 GAFLELAQYPSIFNRTOSNGKIDPSEKITAWAKYFNGNTNILGRGVYVILKNASD 1073
Db 982 GAFLELAQYPSIFELFARKQISTGVPMDSVSKI QWNSAKYFNGNTNILGRGAGYVYLKQOATN 1041
QY 1074 KYFELKGNQ--TYLPMQNTNKAESTGFVNDGNGMTFYSTGYQAKNSFVQDAKGNWYFD 1131
Db 1042 TYFNTSDNKEINFLPKTLLNQDSQVGSFVYDGKGYVYSTGYQAKNTFISEG-DKWYFD 1100
QY 1132 NNGHMYGLQQLNGEVOYFSLNGVOLRESFLENADGSKNYFGLHNGRNGYYSFNDNSK 1191
Db 1101 NNGYMTGAQSIGVNNYVYFSLNGLQRLDALIKNEDGTAYAYGNDGRYENGYQF-MSGV 1159
QY 1192 WRYFDASGVNAVGLKTINGTQYFDQGGQVQKAMITGSDGKKRYPDGSGMNAVNRPAN 1251
Db 1160 WRHFN-NGEMSVGLTVIDQGVYFDEMGYQAKGKFTTADGKIRYPDGSGMNNRFRTE 1218
QY 1252 DKNGDWWYNSDIALGVQTTNGKTYFYGQDGKQIKGKIITD-NGKLKYFLANSGLAR 1310
Db 1219 NEEGKWLJGEDAAVTGSGTNGQHLYFRANGVQVKGFVTDHHRGISYDNGSGDQIR 1278
QY 1311 NIFATDSQNNWYFSGDGVAVTGSQTIAKGLYFASDQKQVKGFSVFTYN-GKVHYHADS 1369

Db 1279 NRVFNAQGWQFYFDNNGYAVTGARTINGQLLYFRANGVQVKGFVTDYRGISYDNGS 1338
QY 1370 GELQVNRFEADKGNWYVYLDNNGEALTGSRINDORVFFTRGKQVKGVDVAYDERRLLVY 1429
Db 1339 GDQIRNRFVRNAQGWQFYFDNNGYAVTGARTINGQHLYFRANGVQVKGFVTDHGRISY 1398

RESULT 14

AAU98035

ID AAU98035 standard; protein; 1475 AA.

XX AAU98035;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant D457N/D571K.

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

FH Misc-difference 457 /note= "Wild-type Asp substituted by Asn"

FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"

FT US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

XX 07-JUN-1995; 95US-00485243.

XX 16-JAN-1998; 98US-00007999.

XX 16-JAN-1998; 98US-00008172.

XX 20-JAN-1998; 98US-00009620.

XX 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
XX B polypeptide having changes at position from 1448V, D457N, D567T,
XX K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
XX I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
XX GTF D polypeptide having changes at positions from T589D, T589E, N471D,
XX N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
XX GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
XX complementary polynucleotide, a ribonucleic acid sequence encoding the
XX GTF mutant, an expression cassette comprising the polynucleotide operably
XX linked to a promoter, a vector comprising the expression cassette, host
XX cell introduced with the vector, a transgenic plant comprising the
XX vector, a seed or tuber from the transgenic plant, a paper sizing and/or
XX coating composition comprising a glucan produced in a plant transformed
XX with a gene encoding the mutant GTF, wild type or, starch, a latex,
XX thermoplastic molecule or their combinations or glucan and starch where
XX the glucan is produced in the amyloplast and/or vacuole or a maize line

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 31.3088 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-36
Perfect score: 7462
Sequence: 1 METKERYKMHKGVHVTVA.....EGQVKGDVAYDERLLVYR 1430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7462	100.0	1430	2	US-09-008-172-2
2	7462	100.0	1430	2	US-09-210-361-6
3	7462	100.0	1430	2	US-09-740-274-6
4	3817.5	51.2	1475	2	US-09-007-999-2
5	3817.5	51.2	1475	2	US-09-210-361-2
6	3817.5	51.2	1475	2	US-09-740-274-2
7	3788.5	50.8	1577	1	US-08-793-824-2
8	3591.5	48.1	1375	2	US-09-210-361-4
9	3591.5	48.1	1375	2	US-09-740-274-4
10	2808	37.6	545	2	US-09-604-957-4
11	2780	37.3	545	2	US-09-995-749A-10
12	2699.5	36.2	2057	2	US-09-499-203-2
13	2446	32.8	1278	2	US-09-604-957-3
14	2446	32.8	1781	2	US-09-995-749A-2
15	1534	20.6	523	2	US-09-604-957-5
16	1532.5	20.5	522	2	US-09-995-749A-11
17	1400	18.8	535	2	US-09-604-957-7
18	1400	18.8	535	2	US-09-995-749A-13
19	1308.5	17.5	584	2	US-09-604-957-6
20	1306.5	17.5	584	2	US-09-995-749A-12
21	595.5	8.0	349	2	US-09-009-620-2
22	453	6.1	2710	1	US-08-480-604A-6
23	453	6.1	2710	1	US-08-405-496A-6
24	453	6.1	2710	2	US-08-915-136-6
25	453	6.1	2710	2	US-08-957-310-6
26	453	6.1	2710	2	US-10-011-366-6
27	453	6.1	2710	2	US-09-084-517-6

28	390.5	5.2	1231	2	US-08-714-741-41	Sequence 41, Appl
29	387	5.2	2366	1	US-08-480-604A-10	Sequence 10, Appl
30	387	5.2	2366	1	US-08-405-496A-10	Sequence 10, Appl
31	387	5.2	2366	2	US-08-915-136-10	Sequence 10, Appl
32	387	5.2	2366	2	US-08-957-310-10	Sequence 10, Appl
33	387	5.2	2366	2	US-10-011-366-10	Sequence 10, Appl
34	387	5.2	2366	2	US-09-084-517-10	Sequence 10, Appl
35	363	4.9	811	1	US-08-480-604A-7	Sequence 7, Appl
36	363	4.9	811	1	US-08-405-496A-7	Sequence 7, Appl
37	363	4.9	811	2	US-08-915-136-7	Sequence 7, Appl
38	363	4.9	811	2	US-08-957-310-7	Sequence 7, Appl
39	363	4.9	811	2	US-10-011-366-7	Sequence 7, Appl
40	363	4.9	811	2	US-09-084-517-7	Sequence 7, Appl
41	363	4.9	812	1	US-08-480-604A-29	Sequence 29, Appl
42	363	4.9	812	2	US-08-915-136-29	Sequence 29, Appl
43	363	4.9	812	2	US-09-084-517-29	Sequence 29, Appl
44	359	4.8	866	2	US-09-545-773-2	Sequence 2, Appl
45	359	4.8	866	2	US-10-222-038-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match	100.0%	Score 7462;	DB 2;	Length 1430;
Best Local Similarity	100.0%	Pred. No. 0;	Mismatches 0;	Indels 0; Gaps 0;
Matches 1430;	Conservative	0;		
QY	1	METKERYKMHKGVHVTVA	VASGLITLGTTLGSSVSAETEQOTS	DKVVTQKSEDDKAA 60
Db	1	METKERYKMHKGVHVTVA	VASGLITLGTTLGSSVSAETEQOTS	DKVVTQKSEDDKAA 60
QY	61	SESSOTDAPKTKQAQTEQ	TAQSQANVADTSTSIKTPSPQNIIT	TAQNSDDKTVTNTKSE 120
Db	61	SESSOTDAPKTKQAQTEQ	TAQSQANVADTSTSIKTPSPQNIIT	TAQNSDDKTVTNTKSE 120
QY	121	EAOTSEERTKQSEEAQT	TASSOALTOAKAELTKQRTAAQENK	NVPDLAAIENVKQIDCK 180
Db	121	EAOTSEERTKQSEEAQT	TASSOALTOAKAELTKQRTAAQENK	NVPDLAAIENVKQIDCK 180
QY	181	YYYIGSDQPKKNFALT	VNNKVLVPDKNTGALTDTSTQ	QFKGLTKLNNDYTPHNOIVNF 240
Db	181	YYYIGSDQPKKNFALT	VNNKVLVPDKNTGALTDTSTQ	QFKGLTKLNNDYTPHNOIVNF 240
QY	241	ENTSLTIDTNYVADSWYR	PKDILKNGKWTATSESDLRPLMS	WWPKQTOIAYLNNYN 300
Db	241	ENTSLTIDTNYVADSWYR	PKDILKNGKWTATSESDLRPLMS	WWPKQTOIAYLNNYN 300
QY	301	QOGLGTGENYVADSSQ	ESLNLAQTVQKIETKISQTOQT	OWLRDIINSFVKTPQNNWQ 360
Db	301	QOGLGTGENYVADSSQ	ESLNLAQTVQKIETKISQTOQT	OWLRDIINSFVKTPQNNWQ 360
QY	361	TESDTSAGEKDLQGG	ALLYSNSDKTAVANSYRLNRT	PTTSQTCGPKYFEDSSGGYDF 420

Db 361 TESDTISAGEKHLQGGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDVADVNADLLQIAD 480
Db 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDVADVNADLLQIAD 480
QY 481 YLKAHYGVDSKSKNAIHLISILEANSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
Db 481 YLKAHYGVDSKSKNAIHLISILEANSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
QY 541 KDAASKNEIRSGLEPVTITNSLNRSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
Db 541 KDAASKNEIRSGLEPVTITNSLNRSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPTAYALMLSNKDSITRLYYGDMYS 660
Db 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPTAYALMLSNKDSITRLYYGDMYS 660
QY 661 DGOYQWATKSPYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
Db 661 DGOYQWATKSPYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQSEATKTQGMVITSNPESLKLQNDKVIIVNMGAAHKNOEYRPLLLTTKQGL 780
Db 721 GANEATDQSEATKTQGMVITSNPESLKLQNDKVIIVNMGAAHKNOEYRPLLLTTKQGL 780
QY 781 TSYTSDAAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
Db 781 TSYTSDAAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLAVVVPVGASDNQDVRVAAS 840
QY 841 NKANATGOVYESSALDSQIIVEGFSNPODFVTIKDSDYTNKKIAQNVQLFKSGVTSPEM 900
Db 841 NKANATGOVYESSALDSQIIVEGFSNPODFVTIKDSDYTNKKIAQNVQLFKSGVTSPEM 900
QY 901 APOVSSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSCQDMINAVKALHKSIGIOVIAD 960
Db 901 APOVSSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSCQDMINAVKALHKSIGIOVIAD 960
QY 961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFLSL 1020
Db 961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFLSL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWEKAKYFNGNTNIIILGRGVGYVLKDNASDKYFELKG 1080
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWEKAKYFNGNTNIIILGRGVGYVLKDNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYIFDNNGHMVYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYIFDNNGHMVYGL 1140
QY 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGLHGNRYSNGYYSFDNDSKWRYPDASGV 1200
Db 1141 QQLNGEVQYFLSNGVQLRESFLENADGSKNYFGLHGNRYSNGYYSFDNDSKWRYPDASGV 1200
QY 1201 MAVGLKTINGNTQYFPDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
Db 1201 MAVGLKTINGNTQYFPDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
QY 1261 NSDGIALGVQTINGNTYYPGDGKQIKGKITDNGKLKYFLANSSELARNIPATDSQNN 1320
Db 1261 NSDGIALGVQTINGNTYYPGDGKQIKGKITDNGKLKYFLANSSELARNIPATDSQNN 1320
QY 1321 WYFFGSDGVAVTGSQTIAGKGLYFASDGKQVKGSVFTYNGKVHYHHADSGELQVNRFEAD 1380
Db 1321 WYFFGSDGVAVTGSQTIAGKGLYFASDGKQVKGSVFTYNGKVHYHHADSGELQVNRFEAD 1380
QY 1381 KQGNWYILDSNGEALTGSQRINDQRFVFTREGKQVKGVDVAYDERRLLVYR 1430
Db 1381 KQGNWYILDSNGEALTGSQRINDQRFVFTREGKQVKGVDVAYDERRLLVYR 1430

US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match 100.0%; Score 7462; DB 2; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 METKRYKMHKVKHWTVAVASGLITLTGTTTGGSSVSAETEQSDSKVVTQKSEDDKAA 60
Db 1 METKRYKMHKVKHWTVAVASGLITLTGTTTGGSSVSAETEQSDSKVVTQKSEDDKAA 60
QY 61 SSSQTDAPKTKQAOTEQTAOSQANVADTSTSIKETPSQNIITQANSDDKTVTNPKSE 120
Db 61 SSSQTDAPKTKQAOTEQTAOSQANVADTSTSIKETPSQNIITQANSDDKTVTNPKSE 120
QY 121 EAQTSEERTKQSEEAQTTASSQALTKQAELTKQRTAAQENKPNVDLAAIPNVKOIDGK 180
Db 121 EAQTSEERTKQSEEAQTTASSQALTKQAELTKQRTAAQENKPNVDLAAIPNVKOIDGK 180
QY 181 YYYIGSDGQPKNFALTNNKVLVYFDKNTGALTDTTSQYQFKQGLTKLNNDYTPHNQIVNF 240
Db 181 YYYIGSDGQPKNFALTNNKVLVYFDKNTGALTDTTSQYQFKQGLTKLNNDYTPHNQIVNF 240
QY 241 ENTSLETIDNYVTADSWTRPKDILKNGKTTWTASSSDLRPLLMWWPDKQTQIAYLNYN 300
Db 241 ENTSLETIDNYVTADSWTRPKDILKNGKTTWTASSSDLRPLLMWWPDKQTQIAYLNYN 300
QY 301 QQGLGTGENYTDSSQESLNLAQTVQVKIETKISQTOQTQWLRLDIINSFVKTPQPNNSQ 360
Db 301 QQGLGTGENYTDSSQESLNLAQTVQVKIETKISQTOQTQWLRLDIINSFVKTPQPNNSQ 360
QY 361 TESDTSAGEKHLQGGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
Db 361 TESDTSAGEKHLQGGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDVADVNADLLQIAD 480
Db 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDVADVNADLLQIAD 480
QY 481 YLKAHYGVDSKSKNAIHLISILEANSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
Db 481 YLKAHYGVDSKSKNAIHLISILEANSNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
QY 541 KDAASKNEIRSGLEPVTITNSLNRSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
Db 541 KDAASKNEIRSGLEPVTITNSLNRSAEGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600

601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
661 DDGOYMATKSPYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
661 DDGOYMATKSPYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
721 GANEATDQGSSEATKTQGMVITSNPSLKLNDKQVIVNMGAAHKNQYERPLLLTTKQGL 780
721 GANEATDQGSSEATKTQGMVITSNPSLKLNDKQVIVNMGAAHKNQYERPLLLTTKQGL 780
781 TSYTSDAAKSLYRKTNDEKELVFDASDIQYLNPOVSGYLAVVVPVPGASDNQDVRVAAS 840
781 TSYTSDAAKSLYRKTNDEKELVFDASDIQYLNPOVSGYLAVVVPVPGASDNQDVRVAAS 840
841 NKANATGOVYESSALDSQLIYEGFSNFPQDFTVKDSYTNKKIAQNVLPKSGVTSFEM 900
841 NKANATGOVYESSALDSQLIYEGFSNFPQDFTVKDSYTNKKIAQNVLPKSGVTSFEM 900
901 APOVYSEDSGSLDSIIQNGYAFEDRYDLAMSKNNKYGSODMINAVKALHKSIGIOVIAD 960
901 APOVYSEDSGSLDSIIQNGYAFEDRYDLAMSKNNKYGSODMINAVKALHKSIGIOVIAD 960
961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
961 WVPDQIYNLPKGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
1021 AAKYPSIFNRTQISNGKKIDPSEKITAWAKAYFNGTNIILGRGVYVLDKONASDKYFELKG 1080
1021 AAKYPSIFNRTQISNGKKIDPSEKITAWAKAYFNGTNIILGRGVYVLDKONASDKYFELKG 1080
1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMYGL 1140
1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMYGL 1140
1141 QQLNGEVQYFSLNGVQLRESFLENADGSKNYFGLHGNRSNGYSPDNDSKWRYFDASGV 1200
1141 QQLNGEVQYFSLNGVQLRESFLENADGSKNYFGLHGNRSNGYSPDNDSKWRYFDASGV 1200
1201 MAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
1201 MAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
1261 NSDGIALVGVQTINGKTYIFGQDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
1261 NSDGIALVGVQTINGKTYIFGQDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKGVSFVYNGKVHYHADSGELQVNRFEAD 1380
1321 WYFSGDGVAVTGSQTIAGKLYFASDGKQVKGVSFVYNGKVHYHADSGELQVNRFEAD 1380
1381 KQGNWYILDSNGEALTGSRINDQRVFFTRGQVKGVDVAYDERLLVYR 1430
1381 KQGNWYILDSNGEALTGSRINDQRVFFTRGQVKGVDVAYDERLLVYR 1430

RESULT 3
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CND
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704

; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match 100.0%; Score 7462; DB 2; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 METKRYKHKVKKHVVTVAVASGLITLGTTLGSSVSAETEQTSDKVVTKSEDDKAA 60
Db 1 METKRYKHKVKKHVVTVAVASGLITLGTTLGSSVSAETEQTSDKVVTKSEDDKAA 60
Qy 61 SESSQTDAPKTQAOEQTOAQOANVADTSITKETPSONITTOQANSDDKVTWTKSE 120
Db 61 SESSQTDAPKTQAOEQTOAQOANVADTSITKETPSONITTOQANSDDKVTWTKSE 120
Qy 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQAKENKPNVDLAAIPNVKQIDGK 180
Db 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQAKENKPNVDLAAIPNVKQIDGK 180
Qy 181 YYYIGSDGPKKNEFALTVNNKVLVFDKNTGALTDSQYQFKOGLTKLNDYTHNQIVNF 240
Db 181 YYYIGSDGPKKNEFALTVNNKVLVFDKNTGALTDSQYQFKOGLTKLNDYTHNQIVNF 240
Qy 241 ENTSLETIDNYVTADSWYRKPDKILKNGKWTWASSSDLRPLLMWMPDKQTOIAYLNYN 300
Db 241 ENTSLETIDNYVTADSWYRKPDKILKNGKWTWASSSDLRPLLMWMPDKQTOIAYLNYN 300
Qy 301 QOGLGTGENYTADSSQESLNLAQAQTVQVKIETKISQTOQWMLRDIINSFVKTPQNNWSQ 360
Db 301 QOGLGTGENYTADSSQESLNLAQAQTVQVKIETKISQTOQWMLRDIINSFVKTPQNNWSQ 360
Qy 361 TESDTSAGEKDLQCGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
Db 361 TESDTSAGEKDLQCGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
Qy 421 LLANDIDNSNPVQAEQLNWLHLYLWNGYSIVANDPEANFDGVVRVDAVDNVDNADLLQIASD 480
Db 421 LLANDIDNSNPVQAEQLNWLHLYLWNGYSIVANDPEANFDGVVRVDAVDNVDNADLLQIASD 480
Qy 481 YLKAHYGVDKSEKNAINHLSTLEAMSDNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
Db 481 YLKAHYGVDKSEKNAINHLSTLEAMSDNDPQYNKDTKGAQLPIDNKLRLSLLYALTRPLE 540
Qy 541 KQASNKNEIRSGLEPVTNSLNRSAGKNSERMANYIPIRAHDSVQTVIAKIIKAQIN 600
Db 541 KQASNKNEIRSGLEPVTNSLNRSAGKNSERMANYIPIRAHDSVQTVIAKIIKAQIN 600
Qy 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
Db 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKKTOSNIPTAYALMLSNKDSITRLYYGDMYS 660
Qy 661 DDGOYMATKSPYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
Db 661 DDGOYMATKSPYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
Qy 721 GANEATDQGSSEATKTQGMVITSNPSLKLNDKQVIVNMGAAHKNQYERPLLLTTKQGL 780
Db 721 GANEATDQGSSEATKTQGMVITSNPSLKLNDKQVIVNMGAAHKNQYERPLLLTTKQGL 780


```
QY 1132 NNGHMYGLQOLNGEVOYFSLNGVQLRESFLENADSGKNYFGLHGNRYSNGYYSFNDK 1191
DB 1101 NNGYMYTGAOSINGVNYFSLNGLQLRDLAKNEEDGTIAYYNGDRYENGYQF-MSGV 1159
QY 1192 WRVFDASGVMAVGLKKTINGNTQYFDODGYOVKGAWITGSGKKRYFDDGSGMVAVRAN 1251
DB 1160 WRHFN-NGEMSUGLTVIDGQVQFDEWGYQAKGFVTTADGKIRYFDKSGMNYRNFIE 1218
QY 1252 DKNGDWYLNDSGIALGVQTINGKTYYPQDQKQIKGKIITD-NGKLYFLANSSELAR 1310
DB 1219 NEEGKWLXGCEGAAVTSQTINGQHLFRANGVQVKGFEVTDHGRISYDNGSDQIR 1278
QY 1311 NIFATDSQNNWYFSGDGVAVTSGQTIAGKLYFASDGKQVKSFTYTN-GKVHYTHADS 1369
DB 1279 NRVFNAQGWYFDDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYDNGS 1338
QY 1370 GELQVRFEADKGNWYILDSNEALTGSDINQDQVFFTRREGQVKGVDVAYDERLLVY 1429
DB 1339 GGOIRNFRVNAQGWYFDDNNGYAVTGARTINGQHLFRANGVQVKGFEVTDHGRISY 1398

RESULT 5
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 035/CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PR1
; ORGANISM: Streptococcus mutans
US-09-210-361-2

Query Match 51.2%; Score 3817.5; DB 2; Length 1475;
Best Local Similarity 52.9%; Pred. No. 1.4e-221;
Matches 762; Conservative 230; Mismatches 395; Indels 53; Gaps 22;

QY 1 METKRYKMHVKCHWTVAVASGLITLCTTTLGSSVSAETEQOTSQDVVTKQSDKAA 60
DB 1 MDKRYKLRKVKCHWTVAVASAVWTL-TTLGGVLVRADSNESK-----QISDNTS 54
QY 61 SSSQTDAPKTKQAQTEQTAQSQAN-VADTSTSTIKETPSQNIITQANSDDKTVTNKS 119
DB 55 VVTANESNVITEATSQEAASQTNHTVTTSSTSTSVNPKVSVNPNVTGSETASNGEK 114
QY 120 EEAQTSERTKQSEAOQTASSQALTQAKAELTKORQTAQENKAPVDLAALPNVKQIDG 179
DB 115 LQNTTTV-DKTSAAANNISQOT-TEADTDVDDSNAA-----NLQILEKLPNVKEIDG 167
QY 180 KYIYIGSDGQPKNFALTVNNKLYFDKNTGALTDS-QYQFKQGLTKLNDYTHNQIV 238
DB 168 KYIYDNGKVRNFTLADGKILHFDG-TGAYTDTSDITVNDKQIVTFRSNDLYKKYNQY 226
QY 239 NFENTSLETIDNYVTADSWYRPKDLKNGKWTWATSESDLRPLLSWPDQQTQIAYLNY 298
```

```
DB 227 DRSAQFEHVHYLTAEWYRPKYILKDGKWTQTSTEKDFRPLLTWMPDQETQRYVNY 286
QY 299 MNOQGLGTGENTTADSSQESLNLAOTVOVKIETKISQTOQOTOWLRDINSFVKTOPNWN 358
DB 287 MNAQ-LGIINKTYDDTSNQLNLAATAIQAIEAKITLTKNTDNLWRTQSISAFVKTOSAWN 345
QY 359 SOTESPTSAGEKDLGGALLYSNSDK-TAYANSDYRLLRNRTPTSTQTK--PKYFEDNSS 415
DB 346 SDSEKFPD---DHLQNGAVLYDNECKLTPYANSNYRILNRTPTNQTGKDPRTADNTI 401
QY 416 GGYDFLLANDIDNSNPVQABQLNWLHLYLMYGSIVANDPEANFDGVRVDVADVNVADLL 475
DB 402 GGYEFLANDVDSNPNVQAEQLNWLHFLMNFENIYANDPDANFDSIRVDADVNDVADLL 461
QY 476 QIASDYLLKAHYGVDSKNAIHNLSILBAWSNDPOYNKDTKGAQLPINKLRLSLLYAL 535
DB 462 QIAGDYLLKAAKGIHKNDKAANDHLSILEAWSNDTPYLHDDGDNMINKLRLSLFL 521
QY 536 TRPLEKDNASNKNRISGLEPVIITNSLNRSABGKNSERMANYIFIRAHDSVQVTIAKII 595
DB 522 AKPLNQ-----RSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLADI 573
QY 596 KQAINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPYALMLSNKDSITRLYY 655
DB 574 KAEINPNVVGYSFTMEIEIKKAFIYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVY 633
QY 656 GMYSDGQYMATKSPYDAIDTLKARIKYAAGQDMKITVYEGDKSHMDWDYTGVLIS 715
DB 634 GDMFTDDGQYMAKHTINYEAIETLLKARIKYVSGGQMRNQVGNSE-----IITS 684
QY 716 VRYGTGANEATDQGSSEATKQGMVITSNPSILKNDKVIWNGMAAHKQNYRPLLT 775
DB 685 VRYGKALKATDTGRTTTSVAVLEGNNPSLRKASDRVVVWNGMAAHKQNYRPLLT 744
QY 776 TKDGLTSTYSDAAAKSLRYKTKNDKGLVFDASDIQYLNPNQVSGYLAVMVPGASQNDV 835
DB 745 TDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGWVPVGA--LI 801
QY 836 RVAASNKANATQGV--YESSALDSQLIYEGFSNQDFVTQSDSYNTKKIAQNVQLKSW 893
DB 802 KMFALRLARPHQOMASVHQNAALDSRVMPFESNFQAFATKKEEYVNVVIAKNDVPAEW 861
QY 894 GYTSFEMAPQYVSSDGSFSDSIIONGYAFEDRYDLAMSKNNKYGQQQMINAVKALHS 953
DB 862 GYTFEMAPQYVSSDGSFSDSVIONGYAFTDRYDLGISKPNKYGTADDLVRAIKALHS 921
QY 954 GIQVIADWVPDQIYNLPQKEVVTATRVNDYGBYRKDSSEIKNTLYAAANTKNGKDYQAKY 1013
DB 922 GIKVMADWVPDQMYAPPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKQQAQY 981
QY 1014 GAFLELAAPYSIFNRTQISNGKIDPSEKITANKAKYFNGTNTILGRGVGVYLDKNSD 1073
DB 982 GAFLELAAPYSIFNRTQISNGKIDPSEKITANKAKYFNGTNTILGRGVGVYLDKNSD 1041
QY 1074 KYFELKNGQ--TYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVODAKGNWYFD 1131
DB 1042 TYFNISDNKEINFLPKTLNQDSQVGFSDGKYGYSTSGYQAKNTFISEG-DKWWYFD 1100
QY 1132 NNGHMYGLQOLNGEVOYFSLNGVQLRESFLENADSGKNYFGLHGNRYSNGYYSFNDK 1191
DB 1101 NNGYMYTGAOSINGVNYFSLNGLQLRDLAKNEEDGTIAYYNGDRYENGYQF-MSGV 1159
QY 1192 WRVFDASGVMAVGLKKTINGNTQYFDODGYOVKGAWITGSGKKRYFDDGSGMVAVRAN 1251
DB 1160 WRHFN-NGEMSUGLTVIDGQVQFDEWGYQAKGFVTTADGKIRYFDKSGMNYRNFIE 1218
QY 1252 DKNGDWYLNDSGIALGVQTINGKTYYPQDQKQIKGKIITD-NGKLYFLANSSELAR 1310
DB 1219 NEEGKWLXGCEGAAVTSQTINGQHLFRANGVQVKGFEVTDHGRISYDNGSDQIR 1278
QY 1311 NIFATDSQNNWYFSGDGVAVTSGQTIAGKLYFASDGKQVKSFTYTN-GKVHYTHADS 1369
```

Db 1279 NREVRNAQOQWFFEDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDGN 1338
QY 1370 GELOVRNRPADKQGNWYLDNNGEALTSQORINDORVFTTRGKQVKGVDVADERLLVY 1429
Db 1339 GDQIRNRFVRNAQOQWFFEDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISY 1398

RESULT 6

US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PR1
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 51.2%; Score 3817.5; DB 2; Length 1475;

Best Local Similarity 52.9%; Pred. No. 1.4e-221;

Matches 762; Conservative 230; Mismatches 395; Indels 53; Gaps 22;

QY 1 METKRYKHVKHVVAVASGLITLGTLLGSSVSAETEQTSDKVVTKSEDDKAA 60
Db 1 MDKRYKLRKRVVTVSVASAVMTL-TTLSSGLVKADSNESKS-----QISDNTS 54
QY 61 SSSOTDAPKTQAOEOTQAOQAN-VADTSTSTTKPTPSQNTTTOANSDDKTVTNKS 119
Db 55 VTANEESNVITEATSKQAASSQTNHTVTSSSTSVVNPKEVSNPTVGETASGEK 114
QY 120 EBAOTSERTKSEEAQTASSQALTOAKAELTKQRTAAENKNPVDLAAIPNVKQIDG 179
Db 115 LQNQTTV-DKTSEAAANNISQKT-TEADTVDDSNAA-----NLQLEKLPNVKEIDG 167
QY 180 KYITGSDGQPKNFALTNVKNVLPDKNVTGALTDS-QYQKQGLTKLNDYTPHNQIV 238
Db 168 KYITDNNNGKVRNFTLADGKILHFE-TGAYTDTSIDTNVKNDIVTTRSNLYKKYNQY 226
QY 239 NFENTSLETIDNVVADSVRPKDILKNGKWTASSESDLRPLLMSSWPDKOTQIAYLNY 298
Db 227 DSAQSFVHDHVLTHAESWIRPKYILKQKWTQSTEKDFRPLMTWPDQETQRYVNY 286
QY 299 MNQQLGIGTGENYTADSSQSLNLAQTQVQKIEKTSQOTQWLRLDIINSFVKTOPNWN 358
Db 287 MNAQ-LGINKTYDDTSNQLQNLNAAATIQAKIEAKITTLKNTDWRQTISAFVKTSAWN 345
QY 359 SQTESDTSAGEKHQLOGGALLVNSDK-TAYANSYVRLNRTPTSTQK--PKYFEDNS 415
Db 346 SDSEKFPD----DHLQNGAVLYDNEGKLTTPYANSYRILNRTPTNQTGKDPRTYADNTI 401
QY 416 GGYDFLLANDIDNSNPVQAEQLNHLFNMFGNIYANDPDANPDSIRVDAVDNVDADLL 475

RESULT 7

US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn

QY 1400 RINDQVFFTRREGKQVGVAY 1421
DB 1537 VINGKQLYFDGSGRQVGRVY 1558

RESULT 8

US-09-210-361-4
; Sequence 4, Application US/09210361

; Patent No. 6284479

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Substitutes for Modified Starches and

; FILE REFERENCE: 0357CR

; CURRENT APPLICATION NUMBER: US/09/210,361

; EARLIER FILING DATE: 1998-12-11

; EARLIER FILING DATE: 1998-01-16

; EARLIER FILING DATE: 1995-06-07

; EARLIER FILING DATE: 1998-01-20

; EARLIER FILING DATE: 1995-06-07

; EARLIER FILING DATE: 1998-01-16

; EARLIER FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 1375

; TYPE: PRT

; ORGANISM: streptococcus mutans

US-09-210-361-4

Query Match 48.1%; Score 3591.5; DB 2; Length 1375;

Best Local Similarity 51.9%; Pred. No. 5.5e-208;

Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 METKRYKMKVKKCHWTVAVASGLITLCTTLTGSSVSASAEBOOTSQSDKVVTKSDDKAA 60
DB 1 MEKVRKFKURKKRVWTVSIAVVTL-TSUGSLVKADS-----TDRQQA 47
QY 61 SSSQTDAPKTQAOEQTAQSQANVADTSTSIKETPSQNIITQANSDDKTIVNTKSE 120
DB 48 VTESQASLVTTSEA-AKETLTATDTSTATSAISOPTATVTDNVSTNQ-----INTTAN 101
QY 121 EAQTSERTQSEEAQTASSQALTOAKA--ELT----- 152
DB 102 TANFVVKPTTTSEQAQNTSDSKIIITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPIK 161
QY 153 -----KQRTAAQENKPNVDLAAIPNVK---QIDGKYYVIGSDGPKQKFNALTNNKVL 203
DB 162 PKIGKLKQSSLSQD-----DIAALGNVKNIRKVGKYYIKEDGLQKNYALNINGKTF 216
QY 204 YFDKNTGALTDSYQFKQGLTKLND-----YTPHQIIVNFENTSLETIDNVVTADSWYR 259
DB 217 FDEE-TGALSNNLTPSKGNIT--NNDNTNSFAQYNQVSTDVANFEHVDHILTAESWYR 273
QY 260 PKDILKNGKTWTASSSDRLPLMSWWPKQTOIAVLNMQOGLGTGENYNTADSSQESL 319
DB 274 PKYILKDGKTWTQSTEKDFRPLMTWPDQETQRQYVNTMAQ-LGIHQTYNTATSPLOL 332
QY 320 NLAAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNNWNSQTESDTSAGEKHQLOGALL 379
DB 333 NLAAQTIQTKIBEKITAENKTNWLRQTIISAFVKTOQSAWNSDSEKPPD-----DHLQKALL 388
QY 380 YSNSDK-TAYANSVYLLNRTPTTSQTK--PKYFEDNSSGGYDFLLANDIDNSNPVQAE 436
DB 389 YSNSKLTQANSYRILNRTPTNQTKGKDPYTDRTTIGGYEFLANDVDNSNPVQAE 448

QY 437 QLNWLHLMYGSIIVANDPEANFGVRVDAVDNVDNADLLOIASDYLKAHYGVYDSEKNAI 496
DB 449 QLNWLHFLMNFYANDPNDPANFDSIRVDADVNDVADLLOIAAGDYLKAAGHKNKDKAAN 508
QY 497 NHLISLEAWSNDPOYNKDTGAQOLPIDNKLRLSLLYALTRPLEKDAASKNRISGLEPV 556
DB 509 DHLISLEAWSYNDTPYLDHDDGDMNMDNRLRLSLLYSLAKPLNQ-----RSGMNP 560
QY 557 ITNSLNNRSAGKNSERMANYIPIRAHDSVQTVIAKIIKAQINPKTDGLTFTLDELKOA 616
DB 561 ITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLIRNIIRTEINPNVVGYSFTTEBKKA 620
QY 617 FKTYNEDMRQAKKKTQSNIPYAYALMLSNKDSITELYYGDMYSDGGOYMATKSPYDAI 676
DB 621 FEIYNKOLLATEKKYTHYNTALSALLTNKSSVPRVYVYDGMPTDGOYMAHKTINYEAI 680
QY 677 DTLLKARIKYAAAGQDMKITVYEGDKSHMDWDYTGVLTSVRYGTGANEATDQSEATKQ 736
DB 681 ETLLKARIKYVSGGQAMRNQOVGNSE-----IITSVRYGKALKATDGTDRTRIS 731
QY 737 GMAVITSNPSLKNQNDKVINMGAHKNQBYRPLLLTKDGLTSYTSDAAKSLYRKT 796
DB 732 GVAVIEGNNPSLRLKASDRVVVNMGAHKNQBYRPLLLTTDNGIKAYHSDQEAAGLVRYT 791
QY 797 NDKGELVFPDASDIQGYLNPQVSGYLAVVVPVCGASDNQDVRVAASNKANATGOVYESSAL 856
DB 792 NDRGELIFTAADIKGYANPQVSGYLGVPVGAADADQDVRVAASTAPSTDGKSVHQAAL 851
QY 857 DSOLIVEGFSNPDQFVTKDSYTNKKIAQNVOLFWSWGVTSFEMAPQYVSSSDGSLDSI 916
DB 852 DSRVMEGFSNPDQFATKEEYTNVVIKNDVKFAEWGVTDFEMAPQYVSSDGSFLDSV 911
QY 917 IQNGYAFEDRYDLAMSNNKNGYSQODMINAVKALHSGIQVIADVWPDQIYNLPGEVVT 976
DB 912 IQNGYAFEDRYDLGI:SKENKYGTADLVKAIKALHSKGI:KVMADWVPDQMYALPEKEVVT 971
QY 977 ATRVNDYGEYRKDSIKNTLYAANTKSNKDYQAKYGGAFSLSELAAKYPSINRTQISNG 1036
DB 972 ATRVDKYGTPVAGSQIKNTLYVVDGKSGKQOQAKYGGAFLELQAKYPELFARKQISTG 1031
QY 1037 KKIDSEKITAMKAKYFNGTNILGRGVYLVKDNASDKYFELKGNQTYLPKQMTN----- 1091
DB 1032 VPMDFSVKLIKQWSAKYFNGTNILGRGAGYLVKQATNTYFSLVSDNTFLPKSLVNPNGHT 1091
QY 1092 KEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAGKNWYFDDNNGHMYGLOQLNGEVOYFL 1151
DB 1092 SSSVTGLVFDGKGYVYSTSGNQAKNAFI-SLGNWYFDDNNGYVMTGAQSIINGANYFL 1150
QY 1152 SNGVOLRESFLENADGSKNYFGLGNRYSGYYSFDNDSKWEYFDASGVMAVGLKTINGN 1211
DB 1151 SNGIQLRNAIYDNGNKNVLSYNGDGRRENGYFLF--GOQWRYFO-NGIMAVGLTRVHGA 1207
QY 1212 TOYFDDQGYQVKGAMITGSDGKKRYFDDGSGNNMVRNFKDNGDWYLLNSDGIALVGVQ 1271
DB 1208 VOYFASGFAQKQPIITADGKLRFRDRDSGNQISNRFVNSKGEWFLFDHNGVATGTV 1267
QY 1272 TINGKTYYPQDGKQIKGIITD-NGKLYFLANGELARNIPATDSQNNWYFSGDGA 1330
DB 1268 TENGQRLYFKPVGQVQAKGFIKANGYLYRYPNSGNEVRNRFVRNSKGEWFLFDHNGIA 1327
QY 1331 VTGSQTIAG 1339
DB 1328 VTGARVWNG 1336

RESULT 9

US-09-740-274-4

; Sequence 4, Application US/09740274

; Patent No. 6465203

; GENERAL INFORMATION:

; APPLICANT: Nichols, Scott E.

; TITLE OF INVENTION: Glucan-containing Compositions and Paper

; FILE REFERENCE: 0357CRD

; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 48.1%; Score 3591.5; DB 2; Length 1375;
Best Local Similarity 51.9%; Pred. No. 5.5e-208;
Matches 721; Conservative 215; Mismatches 350; Indels 103; Gaps 22;

QY 1 METKRYKHKVKKHWTVAVASGLITLGTLLGSSVSAETEQTSDKVVTKSEDDKAA 60
DB 1 MEKRVKFKLKKVKKWTVVSIASAVTL-TSLGSLVKADS-----TDDRQQA 47
QY 61 SSSQTDAPKTKQAQTEQTAQSOANVADTSTITKETPSONITTCQANSDDKTVTNKTSE 120
DB 48 VTESQASLVTTSEA-AKETLTATDTSTATSQPTATVIDNVSTTNQS-----TNTTAN 101
QY 121 EAQTSERTKQSEEAQTTASSQALTOAKA--ELT----- 152
DB 102 TANFVVKPTTTTSEQAKTNSDKIITTSKAVNRLTATGKFPVANNTPAHPKTVTDKIVPIK 161
QY 153 -----KQQTAAQENKPNVDLAAIPNVK---QIDKYYIISGDGPQKQFALTVNKVL 203
DB 162 PKIGKLKQPSLSQD-----DIAALGNVKNIRKVKNGKYYKYKEDGLQKNYALNNGKTF 216
QY 204 YFDKNTGALTDTSQYQKGLTKLND---YTPHNOIVNEFNTSLETIDNVYTDASWR 259
DB 217 FFDE-TGALSNWILPSKKNIT--NNDNTNSFAQYNQVYSTDVANFEHVDHYLTAESWR 273
QY 260 PKDILKNGKWTASSSDLRPLLMSSWPKQTIAYLNTMNOQGLGTGENYTDSSQESL 319
DB 274 PKYILKDGKWTQSTEKDPRLLMTWMPDQETQRQVNYTMAQ-LGIHQTYNTATSPLOL 332
QY 320 NLAQIVQVKIETKISQTOOTQWLRDINSFVKTPQNNNSQTESDTSAGEKHLOGGALL 379
DB 333 NLAQTIQTKIEKIEKTAENKTNLWLRQTIISAFVKTQSANSDSEKPPD-----DHLQKCALL 388
QY 380 YNSDK-TAYANSYVRLNRTPTSTQTK--PKYFEDNSSGGYDFELLANDINDSNPVVQAE 436
DB 389 YNSNKLTSQANSNRIILNRTPTNGKQKPRYTDRTITGGYFELLANDINDSNPVVQAE 448
QY 437 QLNWHLMYNGSIVANDPEANFQVGRVDAVDNVDNADLIQIASDYLKAKHYGVKSEKNAI 496
DB 449 QLNWHLFNMFGNIYANDPANFDSIRVDADVNDVADLLQIAGDYKAAKGHKNDKAA 508
QY 497 NLSLLEAWSNDPOYNKDTGAQPIPNKLRLLLYALTRPLEKDNASKNBSIRSGLEPV 556
DB 509 DLSLLEAWSYNDTPYLHDDGDNMNDNRLRLSLLSLAKPLNQ-----RSGMNP 560
QY 557 ITNSLNNRSABGKNSERNANYIFIRAHSEVQTVIAKIQAINPKTDLGTLTFLDELQOA 616
DB 561 ITNSLNNRDNDAETAAPVSPYSFIRAHSEVQDLIRNIIRTEINPNVVGYSFTTEIKKA 620
QY 617 FKIVNEDMRQAKKKTQSNIPYALMLSNKDSITRLYYGDMYSDDQVMTATKSPYDAI 676

DB 621 FEYNNKOLLATEKXYTHYNTALSYALLLTNKSVPVYYGDMPTDDCQYMAHKTINYEAI 680
QY 677 DTLIKARIKYAAGQDMKITVEYGDKSHMDWDYTVLTSRYGTGANEATDQGEATKTQ 736
DB 681 ETLIKARIKYVSGQAMRNQOVNSE-----IITSVRYKGKALKATDITGDRTRTS 731
QY 737 GMAVITSNPSPSLKLNONDKVINMGAHAKHQRPLLLTTKDGLTSTVSDAAKSLYRKT 796
DB 732 GVAVIEGNNPSSLKASDRVVVNMGAHAKHQAQRPPLLLTTDNGIKATHSQEAAAGLYRYT 791
QY 797 NDKGELVFDASDIQGYLNPQVSGYLAWVPVPGASDNQDVRVAASNKANATGOVYESSAL 856
DB 792 NDRGELIFTAADIKGYANPQVSGYLGWVPVPGAAAQDQDVRVAASTASTDGSKHVQNAAL 851
QY 857 DSQLIYEGESNFODFVTKSDYTNKKIAQNVQLFKSHGVTSPEMAPQVYVSESDGSFLDSI 916
DB 852 DSRVMEFGFSNFQAFATKKEEYTNVVIKRVDFAEWGVTFDEMAPQVYVSTDSGFLDSV 911
QY 917 IQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSIGIOVIADWVPDQIYNLPGEVVT 976
DB 912 IQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSGIKVMADWPDQMYALPEKEVVT 971
QY 977 ATRVNDYGEYRKDSEIKNTLYAANTKSNKDYQAKYGGAFSLSELAARYPSIFNRTQISNG 1036
DB 972 ATRVDKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYGGAFLEELQAKYPFLPARKQISTG 1031
QY 1037 KXIDPEKITTAKAKYFNGTNGILGRGVGVYLVKONADSKYFELKGNQTYLPKQMTN----- 1091
DB 1032 VPMDESVKIKQWSAKYFNGTNGILGRGAGYVVLQQAQNTYFSLVSDNTFLPKSLVNPNGHT 1091
QY 1092 KEASTGFVNDGNGMTFYSTGSAQAKNSFVQDADAKGNMYFDDNGHMYVGLQOLNGEVQYEL 1151
DB 1092 SSVTGLVDFGKGYVYSTGNAQNAFI-SLGNMYFYDDNGYMTVGAOSINGANYIFL 1150
QY 1152 SNGVQLRESFLENADGSKNYFGLHGNRYSNGYYSFNDNSKRYFDSAGVMAVGLKTINGN 1211
DB 1151 SNGIQLRNAIYDNGKNVLSYVNDGRRYENGYYLF--GQWRYFQ-NGIMAVGLTRVHGA 1207
QY 1212 TOYFODGQVQKAMITGSDGKRYFDDGSGMNAVNRFANDKNGDWMYVYVGLNSDIALVGQV 1271
DB 1208 VQYFASGFGQAQGFITTAGDKLRYFDRDSGNQISNRFVRNSKGEWFLFDHNGVAVTGT 1267
QY 1272 TINGKTYFQDGKQIKKIITD-NGKLYFLANSGLARNIFATDSQNNWYVFGSDGVA 1330
DB 1268 ITNGQLRYPKNGVQAKGEFIRDANGLYRYDPNSGNEVRNRPVRNSKGEWFLFDHNGIA 1327
QY 1331 VTGSQTIAG 1339
DB 1328 VTGARVWNG 1336

RESULT 10
US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans

US-09-604-957-4

Query Match 37.6%; Score 2808; DB 2; Length 545;
Best Local Similarity 100.0%; Pred. No. 2.6e-161;
Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 LLANDIDNSNPVQAEQLNHLHYLMNYGSIIVANDPEANFDGVRVDAVNNADLLQIASD 480
DB 1 LLANDIDNSNPVQAEQLNHLHYLMNYGSIIVANDPEANFDGVRVDAVNNADLLQIASD 60

QY 481 YLKAHYGVDSKNAINHLSILEAWSNDNDPQNKDTKGAQLPIDNKLRLSLYALTRPLE 540
DB 61 YLKAHYGVDSKNAINHLSILEAWSNDNDPQNKDTKGAQLPIDNKLRLSLYALTRPLE 120

QY 541 KDAASKNEIRSGLEPVITNSLNNRSABGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 121 KDAASKNEIRSGLEPVITNSLNNRSABGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 180

QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
DB 181 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 240

QY 661 DGOYMATKSPYYDAIDTLLKARIKYAAGQDMKIITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 241 DGOYMATKSPYYDAIDTLLKARIKYAAGQDMKIITYVEGDKSHMDWDYTGVLTSVRYGT 300

QY 721 GANEATDQSEATKTQGMAVITSNPNLSKLNQNDKVIYVNMGAHKNQOEYRPLLLTTKQGL 780
DB 301 GANEATDQSEATKTQGMAVITSNPNLSKLNQNDKVIYVNMGAHKNQOEYRPLLLTTKQGL 360

QY 781 TSYSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVPVGSADNDQVRVAAS 840
DB 361 TSYSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVPVGSADNDQVRVAAS 420

QY 841 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVLPKSWGVTSPFEM 900
DB 421 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVLPKSWGVTSPFEM 480

QY 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSIGIOVIAD 960
DB 481 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSIGIOVIAD 540

QY 961 WVPDQ 965
DB 541 WVPDQ 545

RESULT 11

US-09-995-749A-10

Sequence 10, Application US/09995749A
Patent No. 6867026

GENERAL INFORMATION:
APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
APPLICANT: DIJKHUIZEN, LUBBERT
APPLICANT: RAHAOUI, HAKIM
APPLICANT: LEER, ROBERT-JAN
TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
FILE REFERENCE: BO43388-CIP
CURRENT APPLICATION NUMBER: US/09/995, 749A
CURRENT FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: 09/604,957
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: EPO 00201871.1
PRIOR FILING DATE: 2000-05-25
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 545
TYPE: PRT
ORGANISM: Streptococcus mutans

US-09-995-749A-10

Query Match 37.3%; Score 2780; DB 2; Length 545;

Best Local Similarity 99.6%; Pred. No. 1.3e-159;
Matches 544; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 421 LLANDIDNSNPVQAEQLNHLHYLMNYGSIIVANDPEANFDGVRVDAVNNADLLQIASD 480
DB 1 LLANDIDNSNPVQAEQLNHLHYLMNYGSIIVANDPEANFDGVRVDAVNNADLLQIASD 60

QY 481 YLKAHYGVDSKNAINHLSILEAWSNDNDPQNKDTKGAQLPIDNKLRLSLYALTRPLE 540
DB 61 YLKAHYGVDSKNAINHLSILEAWSNDNDPQNKDTKGAQLPIDNKLRLSLYALTRPLE 120

QY 541 KDAASKNEIRSGLEPVITNSLNNRSABGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
DB 121 KDAASKNEIRSGLEPVITNSLNNRSABGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 180

QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 660
DB 181 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNIPATAYALMLSNKDSITRLYYGDMYS 240

QY 661 DGOYMATKSPYYDAIDTLLKARIKYAAGQDMKIITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 241 DGOYMATKSPYYDAIDTLLKARIKYAAGQDMKIITYVEGDKSHMDWDYTGVLTSVRYGT 300

QY 721 GANEATDQSEATKTQGMAVITSNPNLSKLNQNDKVIYVNMGAHKNQOEYRPLLLTTKQGL 780
DB 301 GANEATDQSEATKTQGMAVITSNPNLSKLNQNDKVIYVNMGAHKNQOEYRPLLLTTKQGL 360

QY 781 TSYSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVPVGSADNDQVRVAAS 839
DB 361 TSYSDDAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVPVGSADNDQVRVAAS 419

QY 840 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVLPKSWGVTSPFEM 899
DB 420 NKANATGOVYESSALDSQLIYEGFSNFODFTVKDSYTNKKIAQNVLPKSWGVTSPFEM 479

QY 900 MAPQVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSIGIOVIAD 959
DB 480 MAPQVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSIGIOVIAD 539

QY 960 DWVPDQ 965
DB 540 DWVPDQ 545

RESULT 12

US-09-499-203-2

Sequence 2, Application US/09499203
Patent No. 6570065

GENERAL INFORMATION:
APPLICANT: KOSSMANN, Jens
APPLICANT: WELSH, Thomas
APPLICANT: QUANZ, Martin
APPLICANT: KNUTH, Karola
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
FILE REFERENCE: 147-196P
CURRENT APPLICATION NUMBER: US/09/499,203
CURRENT FILING DATE: 2000-02-08
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2057
TYPE: PRT
ORGANISM: Leuconostoc mesenteroides

US-09-499-203-2

Query Match 36.2%; Score 2699.5; DB 2; Length 2057;
Best Local Similarity 38.1%; Pred. No. 6.2e-154;
Matches 651; Conservative 221; Mismatches 493; Indels 343; Gaps 45;

QY 2 ETKRRYKMHKVKHVVAVASGLITLGTTLTGSSVSAETEQTSDKVV----TQKSEDD 57
DB 5 ETVTRKKLYKSGKVMVAATAFAVLGVSTVT---TVHADTNSNVAVKQINNTGTNDGSEK 61

QY 58 KAASQSOTDA-----PKTKQ--AQTEOTQAQSQANVA-----DTS--- 91
Db 62 KVPVSTNNDSLKQGTGDFWYDSDGNRVQKTNQIILLTABQLKXNNEKNSLVISDDTSK 121
QY 92 --TSITKETPSQ-----ITTOANSDDKT----- 113
Db 122 DDENISKQTKIANQQTVDATAKGLTNSLSDPITGCHYENHNGYFVVIDASGKQVTLQNI 181
QY 114 -----VNTKSEEAQTSERTKQSE 133
Db 182 DGNLQYFDDNGYQVKGSPRVNGKHIYFDSVTGKASSNVDIVNGKAQGYDAQGNQLKXSV 241
QY 134 EAQTTASS-----QALTO----- 146
Db 242 VADSSQTYFFGNGOPLGLQTIIDGNLQYFNGQGVQIKGFGQVNNKRIYFAPNTGNVAV 301
QY 147 AKAEALTQKQTAQAENKNPV-----DLAA-----IPNVKQIDKVKYVIGSDQ 189
Db 302 ANTEILINGKLOGRDANGNQVKNAFSKDVAGNTFPDPANGVMLTGLQTSKTYLDEQGH 361
QY 190 PKKNFALTNNKVLFPDKNTGALTDTDSQYQFKQGLFKLANDYTPHNOIVNFENTSLETID 249
Db 362 LRKNVAGTFNNOFMYPDADTGAKTAIEYQFQDGLVSQSNENTPHNAAKSYDKSSPENVD 421
QY 250 NYVTADSWYRPKDIILKNGKWTATSSSDDLRLPILMSWPKQTOIAYLNVNQOGLGTGEN 309
Db 422 GYLTDATWYRPTDILKNGDGTASTETDMRPLMTWMPDKQTQANYLNFMSKGLGITT 481
QY 310 YTADSSQESLNLAQOTVQVKIETKSOTOTOWLRDIINSFKVOTPNWNSQTESDTSAGE 369
Db 482 YTAATSQKTLNDAAFIQTAIEQOISLKSSTEWLRDAIDSFKVTQANWKNQTEDEAFDG- 540
QY 370 KDHLOGGALLY--SNSDKTAYANS--PYRLNRTPTSTQTKPKYFEDNSSG--GYDFLLANDI 426
Db 541 LOWLQGGFLAYQDDSHRTPNTDGSNNRKLGRQPINIGDS----KOTDQKSGSEFLANDI 596
QY 427 DSNPNVQAEQLNWLHYLNMYSIIVANDPEANPDGVRVDVANDVNADLLOIASDYLVKAHY 486
Db 597 DSNPNVQAEQLNWLHYLNMYSITGNNDNANFDGIRVDVANDVADLKIAGDYFKALY 656
QY 487 GYDSEKKAINHLSLEAWSDDNDPOVNDTKGAQLPID-----NKLK----- 528
Db 657 GTDKSDANANKHLSLEAWNGKDPQVNVQOQNAQTMQTYVTSQFGNSLTHGANNRNSNM 716
QY 529 --LSLLYALTREPLEKASGNKNEIRSGLEPVTISSLNRSAGKSNRMANIYFIRAHDS 586
Db 717 YFLDTGYYLNGDLNKKIVDKNRPNSS-----TLNVRIANSQDTKVPINYSFVRAHYD 769
QY 587 VQTUTAK-----IITKAQINPKTDGLTFTLDELKQAFKIYNEDMROAK--KKYTSQNTPT 638
Db 770 AQDPTRKAMIDHGIKNNMQD-----TFTFDQAQGMFEYKQDQENPSPGFKKYNDYNLPS 823
QY 639 AYALMLSNKDSITRLYYGDMYSDDOYMATKSPYYDAITLLKARIKYAAGQDMKIITYV 698
Db 824 AYAMLLTNKQTVPRVYGDMLYLEGQYMEKGTIYNPVSALLKARIKYVSGSQTMAT--- 880
QY 699 EGDKSHMDW--DYTGVLTSVRYGTGA---NEATDQGSSEATKTOGMVITSNPSSLKNO 752
Db 881 --DSSGKDLKQGETDILLTSVRFKGMTSDQTTTQDQNSQDYKNGIIVGNPNLKLNN 938
QY 753 NDKVIVNNGAAHKNQYRPLLTTKDGLTSYTSDAAKSLYKRTKNDKGELVPDAS----- 807
Db 939 DKTITLHMGAHKNQYRALVLSNDSGIDVYSDDKAPT--RTNDNGDLIFHKNTFPVK 996
QY 808 -----DIOGLYNPQVSGYLAWVPVPGASDNQDVR--VAASNKANATQVTVSSSALDS 858
Db 997 QDGTIINTEMKGSNLSLISGLVWVPVPGASDSQDARTVATESSSNDGVSFHSNAALDS 1056
QY 859 QLIYEGFSNFODFVTKDSQDYNKTKIAONVQLFKSMGVTSFEMAPQVSSSDG-----SFL 913
Db 1057 NVIYEGFSNFQAMPTSPQSTNVNVIATKANLFKELGITSFELAPQYRSSGDTNYGGMFL 1116
QY 914 DSIITQNGYAFEDRYDLAMSKNN-----KYGSDQDMINAVKALHKSQIVIADWVPDQIY 967

Db 1117 DSQLNNGYAFTRDYDILGFNKAQDGNPNPTKYGTDDLRNAIEALHKNQMQIADWVPDQIY 1176
QY 968 NLPKGKVVWVATRVNDYGEYRKDSEIKNTLYAAANTKSGKDYQAKYGGAPLSELAAYPSI 1027
Db 1177 ALFPGKVVWVATRVDERGNQLKOTDFVNLIVYANTKSSGVYQAKYGEFLLKUREEYPSL 1236
QY 1028 FNRTQISNKKKIDPSEKITAUKAKYFNGTNIILGRGVGYVLKONASDKYFEL--KGNQTYLP 1086
Db 1237 FKQNOVSTGQPIDASTKIKQWSAKYMNGTNIIILHRGAYVYLVKOWATQYFNIAKTNEVFLP 1296
QY 1087 KQMTNKEASTGFVNDGNGMTFTSYGQAKNSFVQDAKGNWYFFDNGHMYGLO----- 1141
Db 1297 LQLOKNDAQTGFISDASGVKYYSISGYQAKDTPIEDGNGWYFFDKGYMVRSGQGENPI 1356
QY 1142 -----QLNGEVOYFLSNGVQLRESFLENADGSKNYFCHLGNRYSNGYSFNDNSKWR 1193
Db 1357 RIVETSIVNTRNGN--YFPMENGVELKRGFGTDSNGVYFFDDQCKMVRDKYIN--DDANNFY 1414
QY 1194 YFDASGVMAVGLKTINGNT--QYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRPAN- 1251
Db 1415 HLNVDGTMRSRGLFKFEDSDTLQYFASNGVQIKDSYAKDSKGNKYFFDSATGNNDTGAQTW 1474
QY 1252 DKNGDHYLYNSDGIALLVGQV-----IN 1274
Db 1475 DNGYIYITIDSANNTIGVNTDYTAYITSSLRDGLFANAPYGVVTKDQNGDLKWQYIN 1534
QY 1275 GXTYVFGQ-----DGKQIKGIITDNG--KLKYFLANSSELARINIFATDSQNNW- 1321
Db 1535 HTKQYEGQOVQVTRQYTSKGVSWNLITPAGDGLQGRUWVDSRALMTPTPKTMNQISFI 1594
QY 1322 --YFGSDGVAVTGSQTIAGKLYFASDGKQVKGSFVYTKGVHYHADSGELQVNRFEAD 1380
Db 1595 SYANRNDGLFLNAPYQVKGQYL--AGMSNQYKQGVTIAGVANVSGKWSLISFN----- 1647
QY 1381 KQGNWYLYDSNGEAL--TGSQRINDQRVF 1407
Db 1648 --GTQYWIDS--QALNTNFTHDMNQKF 1671
RESULT 13
US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3
Query Match 32.8%; Score 2446; DB 2; Length 1278;
Best Local Similarity 46.7%; Pred. No. 6,1e-139;
Matches 508; Conservative 170; Mismatches 310; Indels 100; Gaps 25;
QY 174 VKQIDGKYYIG-SDQPKKNFALTNNKVLVFDKNTGALTDTDSQYQFKQGLTKLNNDYT 232
Db 235 LQTINGQYYIDPTTQGPQRKFLQSGNNWYFFDSDTGVGTVALBELQAKGVTVSSNEQYR 294
QY 233 PHNOIVNFENTSLETIDNYVTADSWYRPKDIILKNGKWTATSSSDDLRLPILMSWPKQTOI 292

```
Db 295 NGNAA5YDDKSIENVNGYLTDWYRPKQILKOGTWTWDSKETDMRPILMWNTLTQ 354
QY 293 IAYLNNMQOGLGTGEN-----YTADSSOESLNLAQTQVQVKIETKISOTQOTQMLR 344
Db 355 AYLYNNMQHG-----NLLPSALPFENADADPAELNHYSEIVQONIEKRISETGTNDMLR 409
QY 345 DIINSFVKTPQPNWNSQTSBDSAGSKDHLQGGALLYNSDDKTAYANSYRLLNRPPTSQT 404
Db 410 TLMHDFVTNNPMWNKDSNVNFSGIQ--FOGFLKYNSDLTPYANSYRLLGRMPIN-- 465
QY 405 GKPKYFEDNSGGYDFLLANDIDNSNPVVQABQLNWLHYLMNYGSIIVANDPEANFDGVRV 464
Db 466 -----IKDQYRGQBFLLANDIDNSNPVVQABQLNWLHYLLNFGTITANNDOANFDSVRV 520
QY 465 DAVDNVNADLLOIASDYLKAHYGVDSKSEKNAIHLNLSLEAWSDNDPQYNKDTKGALPID 524
Db 521 DAPDNIDADLMNIAQDYFNAAYGMD--SDAVSNKHINILEDWNHADPEYFNKIGNQLTMD 579
QY 525 NKRLSLLYALTRPLEKQASNKNEIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAH 584
Db 580 DTIKNSLNHGLS-----DATN----RWGLDAIVHQSADRENNTENSTENVIPNYSFVRAD 630
QY 585 SEVQTVIAKIIKAQINPKTDGLTFTLDELQAFKIYNEDMRQAKKYTOSNIPTAVALML 644
Db 631 NNSDQIQONAIR-DVTGK--DYHTFTFEDEQKIDAYIQDQNSTVKKNLYNIPASTAILL 688
QY 645 SNKDSITRLYYGDMYSDGOYNATKSPYYDAIDTLIKARIKYAAGQDMKIITYVEGDKSH 704
Db 689 TNKDITPRVYGDLYTDCGOYMEHQTRYDYDTLNLKSEVKYVAGGOSQMTSVGNN-- 747
QY 705 MDWDYTGVLTSVRYGTGANEATDQSEATKTQGMAVITNSNPSLKLQNDKVIIVNMGAAH 764
Db 748 -----ILTSVRYGKGAMTATDGTDETRTQIGVGVVSNTPNLKLGVDNKVVLHMGAAH 800
QY 765 KNQYRPLLTTKDGLTSTSDAAKSLYRKNTDKGELVFDASD-----IQGY 812
Db 801 KNQYRAAVALTTDGVINYSDOGAP--VAMTDENGDLVLSHNLVNGKEEADTAVQGY 858
QY 813 LNPQVSGYLAWVPVGCASDNQDVRVAASNKANATGOVYESSALDSQLITYEGFSNFQDFV 872
Db 859 ANPDVSGYLAWVPVGCASDNQDARTAPSTEKNSGNSAYETNAAPFDSNVIFEAFSNVYTP 918
QY 873 TKSDYTNKIIAQNVLFKSWGVTGFEMAPQVSVSDESGFLDSIIQNGYAFEDRYDLAMS 932
Db 919 TKESERANVRIAQNADFFASLGTFSFEMAPQVNSKKDRTFLDSTIDNGYAFTRDYDLGMS 978
QY 933 KKKYGSQODMINAVKALHKSIGIOTADVPOIYNLPKGEVVVTRVNDYGEYKDSFI 992
Db 979 EPNKYGTDEDLRNAIQALHKAQLQVADWVPQIYNLPKGEVATVTRVDDRGVNWKDALI 1038
QY 993 KNTLYAANTKSGKDYQAKYGAFISELAAKYPSIFNRQTQISNGKKIDPSEKITAWKAKY 1052
Db 1039 NNNLYVNTTIGG--EYQKYGAFDLKQKLPELFTKKQVSTGVADPSQKITWSAKY 1097
QY 1053 FNGTNILRGVGYVLKONASDKYFELKGNQOT--YLPKQMT--NKEASTGFV--NDNGMT 1106
Db 1098 FNGTNILHRGSGVLKADGG-QYNL-GTTTQKQFLPQLTGKSKQGNFVKGNDGN--YY 1154
QY 1107 FYSTSGYQAKNSFVQDAKGNWYFDDNGHMVYGLQQLN-----GE--VQYFLSNGVLR 1160
Db 1155 FYDLAGNMVKNFTFIEDSVGNWYFFDQDGHQVKNKHFPVDVDSYGEKGTFFLKNKGSFRRG 1214
QY 1161 FLENADGSKNYFCHLGNRYSGNYSFNDNSKWRYPDASGVMAVGLKTINGNTQYF--DQD 1218
Db 1215 LNQT-----DNGTYFDNYGK-----WVRNQTINAGAMIYTLDEN 1249
QY 1219 GYQVKGAW 1226
Db 1250 GKILIRASY 1257
Query Match 32.8%; Score 2446; DB 2; Length 1781;
Best Local Similarity 46.7%; Pred.No. 9.9e-139; Mismatches 310; Indels 100; Gaps 25;
Matches 508; Conservative 170;
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUTI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2
QY 174 VKQIDGKYYIG-SDGQPKQNFALTNNKNVLYFDKNTGALTDTQSVQFKQGLTKLNDYT 232
Db 738 LQTNGQYYIDPTTGQPKRNFLLQSGNNWYFDSDTGVTNALSLELQFAKGTVSSNEQVR 797
QY 233 PHQIVNFPENTSLEITDNYVTADSWYRPKDILKNGKWTWTSSESDLRPLMLSWPDKQTO 292
Db 798 NGNAA5YDDKSIENVNGYLTDWYRPKQILKOGTWTWDSKETDMRPILMWNTLTQ 857
QY 293 IAYLNNMQOGLGTGEN-----YTADSSOESLNLAQTQVQVKIETKISOTQOTQMLR 344
Db 858 AYLYNNMQHG-----NLLPSALPFENADADPAELNHYSEIVQONIEKRISETGTNDMLR 912
QY 345 DIINSFVKTPQPNWNSQTSBDSAGSKDHLQGGALLYNSDDKTAYANSYRLLNRPPTSQT 404
Db 913 TLMHDFVTNNPMWNKDSNVNFSGIQ--FOGFLKYNSDLTPYANSYRLLGRMPIN-- 968
QY 405 GKPKYFEDNSGGYDFLLANDIDNSNPVVQABQLNWLHYLMNYGSIIVANDPEANFDGVRV 464
Db 969 -----IKDQYRGQBFLLANDIDNSNPVVQABQLNWLHYLLNFGTITANNDOANFDSVRV 1023
QY 465 DAVDNVNADLLOIASDYLKAHYGVDSKSEKNAIHLNLSLEAWSDNDPQYNKDTKGALPID 524
Db 1024 DAPDNIDADLMNIAQDYFNAAYGMD--SDAVSNKHINILEDWNHADPEYFNKIGNPOLTMD 1082
QY 525 NKRLSLLYALTRPLEKQASNKNEIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAH 584
Db 1083 DTIKNSLNHGLS-----DATN----RWGLDAIVHQSADRENNTENSTENVIPNYSFVRAD 1133
QY 585 SEVQTVIAKIIKAQINPKTDGLTFTLDELQAFKIYNEDMRQAKKYTOSNIPTAVALML 644
Db 1134 NNSDQIQONAIR-DVTGK--DYHTFTFEDEQKIDAYIQDQNSTVKKNLYNIPASTAILL 1191
QY 645 SNKDSITRLYYGDMYSDGOYNATKSPYYDAIDTLIKARIKYAAGQDMKIITYVEGDKSH 704
Db 1192 TNKDITPRVYGDLYTDCGOYMEHQTRYDYDTLNLKSEVKYVAGGOSQMTSVGNN-- 1250
QY 705 MDWDYTGVLTSVRYGTGANEATDQSEATKTQGMAVITNSNPSLKLQNDKVIIVNMGAAH 764
Db 1251 -----ILTSVRYGKGAMTATDGTDETRTQIGVGVVSNTPNLKLGVDNKVVLHMGAAH 1303
QY 765 KNQYRPLLTTKDGLTSTSDAAKSLYRKNTDKGELVFDASD-----IQGY 812
Db 1304 KNQYRAAVALTTDGVINYSDOGAP--VAMTDENGDLVLSHNLVNGKEEADTAVQGY 1361
QY 813 LNPQVSGYLAWVPVGCASDNQDVRVAASNKANATGOVYESSALDSQLITYEGFSNFQDFV 872
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 127.994 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-36

Perfect score: 7462
Sequence: 1 METKRYKMKHKVKKHWTVA.....EGKQVKGDVAYDERRLLVTR 1430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA_Main:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7462	100.0	1430	3	US-09-740-274-6
2	7462	100.0	1430	4	US-10-383-930-36
3	7462	100.0	1430	5	US-10-797-821-36
4	3817.5	51.2	1475	3	US-09-740-274-2
5	3817.5	51.2	1475	4	US-10-383-930-34
6	3817.5	51.2	1475	5	US-10-797-821-34
7	3615	48.4	1554	4	US-10-383-930-38
8	3615	48.4	1554	5	US-10-797-821-38
9	3591.5	48.1	1375	3	US-09-740-274-4
10	3591.5	48.1	1375	4	US-10-383-930-35
11	3591.5	48.1	1375	5	US-10-797-821-35
12	3575	47.9	1590	4	US-10-383-930-37
13	3575	47.9	1590	5	US-10-797-821-37
14	3502.5	46.9	1518	4	US-10-383-930-40
15	3502.5	46.9	1518	5	US-10-797-821-40
16	3364.5	45.1	1497	5	US-10-484-218-18
17	3174	42.5	1365	4	US-10-383-930-39
18	3174	42.5	1365	5	US-10-797-821-39
19	3067	41.1	1595	3	US-10-484-218-20
20	2780	37.3	545	3	US-09-995-749A-10
21	2699.5	36.2	2057	4	US-10-417-280A-2
22	2460.5	33.0	1777	5	US-10-484-218-12
23	2446	32.8	1781	3	US-09-995-749A-2
24	2380.5	31.9	1771	5	US-10-484-218-14
25	2349	31.5	1006	5	US-10-484-218-22
26	1532.5	20.5	522	3	US-09-995-749A-11
27	1400	18.8	535	3	US-09-995-749A-13

28	1306.5	17.5	584	3	US-09-995-749A-12	Sequence 12, Appl
29	1257.5	16.9	787	5	US-10-484-218-16	Sequence 16, Appl
30	1246	16.7	525	5	US-10-484-218-23	Sequence 23, Appl
31	668.5	9.0	224	5	US-10-484-218-4	Sequence 4, Appl
32	664	8.9	223	5	US-10-484-218-10	Sequence 10, Appl
33	647	8.7	223	5	US-10-484-218-6	Sequence 6, Appl
34	515	6.9	221	5	US-10-484-218-8	Sequence 8, Appl
35	509	6.8	221	5	US-10-484-218-2	Sequence 2, Appl
36	453	6.1	2710	4	US-10-011-366-6	Sequence 6, Appl
37	453	6.1	2710	4	US-10-354-774-6	Sequence 6, Appl
38	453	6.1	2710	4	US-10-271-012-6	Sequence 6, Appl
39	453	6.1	2710	4	US-10-729-122-6	Sequence 6, Appl
40	453	6.1	2710	4	US-10-729-039-6	Sequence 6, Appl
41	453	6.1	2710	5	US-10-729-527-6	Sequence 6, Appl
42	453	6.1	2710	5	US-10-727-898-6	Sequence 6, Appl
43	453	6.1	2710	5	US-10-728-696-6	Sequence 6, Appl
44	453	6.1	2710	6	US-11-001-241-6	Sequence 6, Appl
45	387	5.2	2366	4	US-10-011-366-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; TYPE: PRT
; ORGANISM: streptococcus mutans
; US-09-740-274-6

Query Match	100.0%;	Score	7462;	DB	3;	Length	1430;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	1430;	Conservative	0;	Mismatches	0;	Indels	0;
QY	1	METKRYKMKHKVKKHWTVA	ASGLITLGTTLGSSVSAETEQTS	DKVVTQKSEDDKAA	60		
DB	1	METKRYKMKHKVKKHWTVA	ASGLITLGTTLGSSVSAETEQTS	DKVVTQKSEDDKAA	60		
QY	61	SSSQTDAPKTQKQAEQTQ	QAQSNVADTSTSI	TKETPSQNTTQANSDDKTVTN	120		
DB	61	SSSQTDAPKTQKQAEQTQ	QAQSNVADTSTSI	TKETPSQNTTQANSDDKTVTN	120		
QY	121	EAQTSEERTKQSEEAQT	TASSQAALTOAKAELTK	QRTAQENKNPVDLAAI	PNVKIDGK	180	
DB	121	EAQTSEERTKQSEEAQT	TASSQAALTOAKAELTK	QRTAQENKNPVDLAAI	PNVKIDGK	180	
QY	181	YYVIGSDGQPKKNFAL	TNNKVLVFDKNTGALT	DTSOYQFKGLTKLNNDY	TPHNOIVNF	240	
DB	181	YYVIGSDGQPKKNFAL	TNNKVLVFDKNTGALT	DTSOYQFKGLTKLNNDY	TPHNOIVNF	240	

Db 181 YYYIGSDGPKKFNALTANNKVLVYFDKNTGALTDTTSQYQKQGLTKLNNDYTPHNQIVNF 240
Qy 241 ENTSLETIDNYTADSWYRPPKQILKNGKTTWTASSSDLRPLLMSWMPDKQTOIAYLNYN 300
Db 241 ENTSLETIDNYTADSWYRPPKQILKNGKTTWTASSSDLRPLLMSWMPDKQTOIAYLNYN 300
Qy 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTQWLRLDIINSFVKTPQNNNSQ 360
Db 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTQWLRLDIINSFVKTPQNNNSQ 360
Qy 361 TESDTSAGEKDHQGLGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
Db 361 TESDTSAGEKDHQGLGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420
Qy 421 LLANDIDNSNPVQAEQLNWLHYLMNYGSIIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480
Db 421 LLANDIDNSNPVQAEQLNWLHYLMNYGSIIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480
Qy 481 YLKAHYGVDSKSEKNAIHLNLSILEAWSNDPQYNKDTKGAQLPIDNKLRLSLIYALTRPLE 540
Db 481 YLKAHYGVDSKSEKNAIHLNLSILEAWSNDPQYNKDTKGAQLPIDNKLRLSLIYALTRPLE 540
Qy 541 KDAKNEIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
Db 541 KDAKNEIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKAQIN 600
Qy 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPATYALMLSNKDSITRLYYGDMYS 660
Db 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPATYALMLSNKDSITRLYYGDMYS 660
Qy 661 DDQYMATKSPYYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYCT 720
Db 661 DDQYMATKSPYYDAIDTLLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYCT 720
Qy 721 GANEATDQGESEATKTQGMVITSNPDLKLNQNDKVIIVNMGAAHKNQYRPLLLTTKQGL 780
Db 721 GANEATDQGESEATKTQGMVITSNPDLKLNQNDKVIIVNMGAAHKNQYRPLLLTTKQGL 780
Qy 781 TSYTSDAAKSLYRKNTDKGELVFDASDIQGYLNPOVSGYLAVWVPVGASDNQDVRVAAS 840
Db 781 TSYTSDAAKSLYRKNTDKGELVFDASDIQGYLNPOVSGYLAVWVPVGASDNQDVRVAAS 840
Qy 841 NKANATGOVYESSALDSQLIYEGFSNFODFTKDSYTNKKTAAQNVQLFKSNGVTSFEM 900
Db 841 NKANATGOVYESSALDSQLIYEGFSNFODFTKDSYTNKKTAAQNVQLFKSNGVTSFEM 900
Qy 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGIQUIAD 960
Db 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGIQUIAD 960
Qy 961 WVPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAPLSEL 1020
Db 961 WVPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAPLSEL 1020
Qy 1021 AAKYPSIFNRTOISNGKKIDPSKITAWEKAFNGNIIILGRGVYVLKDNASDKYFELKG 1080
Db 1021 AAKYPSIFNRTOISNGKKIDPSKITAWEKAFNGNIIILGRGVYVLKDNASDKYFELKG 1080
Qy 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMVYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMVYGL 1140
Qy 1141 QQLNGEVQFLSNGVQLRESFLENADGSKNYFGHLGNRYSNGYYSFPDNDKWRYPDASGV 1200
Db 1141 QQLNGEVQFLSNGVQLRESFLENADGSKNYFGHLGNRYSNGYYSFPDNDKWRYPDASGV 1200
Qy 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDGSGNMAVNRFANDKNGDWWYL 1260
Db 1201 MAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDGSGNMAVNRFANDKNGDWWYL 1260
Qy 1261 NSDGIALVGQTINGKTYTFGQDGKQIKGKIITDNGKLXYFLANSSELARNIPATDSQNN 1320
Db 1261 NSDGIALVGQTINGKTYTFGQDGKQIKGKIITDNGKLXYFLANSSELARNIPATDSQNN 1320

Qy 1321 WYFEGSDGVAVTGSOTIAGKKLYFASDGKQVKGSVFTYNGKVHYVHADSGELQVNRFEAD 1380
Db 1321 WYFEGSDGVAVTGSOTIAGKKLYFASDGKQVKGSVFTYNGKVHYVHADSGELQVNRFEAD 1380
Qy 1381 KDGNNYYILDSNGEALTGSQRINDQRVFFTRREGKQVKGVDVAYDERLLVYR 1430
Db 1381 KDGNNYYILDSNGEALTGSQRINDQRVFFTRREGKQVKGVDVAYDERLLVYR 1430

RESULT 2

US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRP
; ORGANISM: Streptococcus mutans
US-10-383-930-36

Query Match 100.0%; Score 7462; DB 4; Length 1430;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 METKERYKMHKVKHWHVTAVASGLITLGTTLTGLSSVSAETEQOTSDKVVTKQSEDDKAA 60

Db 1 METKERYKMHKVKHWHVTAVASGLITLGTTLTGLSSVSAETEQOTSDKVVTKQSEDDKAA 60

Qy 61 SSSOTDAPKTKQAOTEQTQAOQANVADTSTSIKTETPSQNTTQAASDDKTVNTKSE 120

Db 61 SSSOTDAPKTKQAOTEQTQAOQANVADTSTSIKTETPSQNTTQAASDDKTVNTKSE 120

Qy 121 EAQTSSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQIDGK 180

Db 121 EAQTSSEERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQIDGK 180

Qy 181 YYYIGSDGPKKFNALTANNKVLVYFDKNTGALTDTTSQYQKQGLTKLNNDYTPHNQIVNF 240

Db 181 YYYIGSDGPKKFNALTANNKVLVYFDKNTGALTDTTSQYQKQGLTKLNNDYTPHNQIVNF 240

Qy 241 ENTSLETIDNYTADSWYRPPKQILKNGKTTWTASSSDLRPLLMSWMPDKQTOIAYLNYN 300

Db 241 ENTSLETIDNYTADSWYRPPKQILKNGKTTWTASSSDLRPLLMSWMPDKQTOIAYLNYN 300

Qy 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTQWLRLDIINSFVKTPQNNNSQ 360

Db 301 QQGLGTGENYTADSSQESLNLAQTQVVKIETKISQTOQTQWLRLDIINSFVKTPQNNNSQ 360

Qy 361 TESDTSAGEKDHQGLGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420

Db 361 TESDTSAGEKDHQGLGALLYNSDKTAYANSYRLLNRTPTSTQTKPKYFEDNSSGGYDF 420

Qy 421 LLANDIDNSNPVQAEQLNWLHYLMNYGSIIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480

Db 421 LLANDIDNSNPVQAEQLNWLHYLMNYGSIIVANDPEANFDGVRVDAVDNVDNADLLQIASD 480

Qy 481 YLKAHYGVDSKSEKNAIHLNLSILEAWSNDPQYNKDTKGAQLPIDNKLRLSLIYALTRPLE 540

Db 481 YLKAHYGVDSKSEKNAIHLNLSILEAWSNDPQYNKDTKGAQLPIDNKLRLSLIYALTRPLE 540

QY 541 KDAASNKNEIRSGLEPVITNSLNRSABGKNSRMANYIFIRAHDSVQTVIAKIIKAAQIN 600
DB 541 KDAASNKNEIRSGLEPVITNSLNRSABGKNSRMANYIFIRAHDSVQTVIAKIIKAAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYVTSQNTPTAVALMLSNKDSITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYVTSQNTPTAVALMLSNKDSITRLYYGDMYS 660
QY 661 DDQGYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDQGYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDOGESEATKQGMVITNSNPSLKLNDKQIVNMGAAHKNQYRPLLLTTKQGL 780
DB 721 GANEATDOGESEATKQGMVITNSNPSLKLNDKQIVNMGAAHKNQYRPLLLTTKQGL 780
QY 781 TSVTSDAASAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLAVVPPVGASDNQDVRVAAS 840
DB 781 TSVTSDAASAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLAVVPPVGASDNQDVRVAAS 840
QY 841 NKANATQGVYESSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQNVQLFKSWGVTSEFM 900
DB 841 NKANATQGVYESSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQNVQLFKSWGVTSEFM 900
QY 901 APOYVSESDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVKALHKSIGIOVIAD 960
DB 901 APOYVSESDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVKALHKSIGIOVIAD 960
QY 961 WPDQIYNLPGEKVTVATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
DB 961 WPDQIYNLPGEKVTVATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAARAKYFNGTNILGRGVYVLDKNASDKYFELKG 1080
DB 1021 AAKYPSIFNRTQISNGKKIDPSEKITAARAKYFNGTNILGRGVYVLDKNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGSYQAKNSFVQDAKGNWYFDNNHVMYGL 1140
DB 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGSYQAKNSFVQDAKGNWYFDNNHVMYGL 1140
QY 1141 QOLNGEVOYFLSNGVOLRESFLENADGSKNYFGLGNRYSNGYYSFDNDSKRYFPDASGV 1200
DB 1141 QOLNGEVOYFLSNGVOLRESFLENADGSKNYFGLGNRYSNGYYSFDNDSKRYFPDASGV 1200
QY 1201 MAVGLKTINGNTQYFQDQGVQVKGAWITGSDGKRYFDDGSGNMAVNRFANDKNGDWYVL 1260
DB 1201 MAVGLKTINGNTQYFQDQGVQVKGAWITGSDGKRYFDDGSGNMAVNRFANDKNGDWYVL 1260
QY 1261 NSDGIALVGQTINGKTYTFYFGDGKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320
DB 1261 NSDGIALVGQTINGKTYTFYFGDGKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320
QY 1321 WYFEGSDVAVTGSQTIAGKLYFASDGQVKGFSFVYNGKVHYHADSSELQVNRFEAD 1380
DB 1321 WYFEGSDVAVTGSQTIAGKLYFASDGQVKGFSFVYNGKVHYHADSSELQVNRFEAD 1380
QY 1381 KQGNWYILDSNGEALTGSRINDQRFVFFTRREGQVKGVDVAYDERLLVYR 1430
DB 1381 KQGNWYILDSNGEALTGSRINDQRFVFFTRREGQVKGVDVAYDERLLVYR 1430

RESULT 3
US-10-797-821-36
; Sequence 36, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930

; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-36

Query Match 100.0%; Score 7462; DB 5; Length 1430;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METKRYKMHKVKHWHVTVAVASGLITLGTTLTGSSVSAETEQTSDKVVTKSEDDKAA 60
DB 1 METKRYKMHKVKHWHVTVAVASGLITLGTTLTGSSVSAETEQTSDKVVTKSEDDKAA 60
QY 61 SSSQTDAPKTKQAQTEQTAQSOANVADTSTSIKTETPSQNTTQANSDDKTVTWKSE 120
DB 61 SSSQTDAPKTKQAQTEQTAQSOANVADTSTSIKTETPSQNTTQANSDDKTVTWKSE 120
QY 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
DB 121 EAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
QY 181 YYYIGSDGPKKFNALTVNNKVLVFDKNTGALTDTSTQYQFKQLTKLNNDYTPHNQIVNF 240
DB 181 YYYIGSDGPKKFNALTVNNKVLVFDKNTGALTDTSTQYQFKQLTKLNNDYTPHNQIVNF 240
QY 241 ENTSLETIDNYTADSWYRPKDILKNGKWTWASSEDLPPLMSWMPDKQTOIAYLNYN 300
DB 241 ENTSLETIDNYTADSWYRPKDILKNGKWTWASSEDLPPLMSWMPDKQTOIAYLNYN 300
QY 301 QOGLGTGENYADSSQESLNLAQTVQVKIETKISQTOQWLRDINSFVKTPQNNNSQ 360
DB 301 QOGLGTGENYADSSQESLNLAQTVQVKIETKISQTOQWLRDINSFVKTPQNNNSQ 360
QY 361 TESDTSAGEKHLQGGALLYNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
DB 361 TESDTSAGEKHLQGGALLYNSDKTAYANSYRLNRTPTSTQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQNLWHLVNYGSIIVANDPEANFDGVRVADVNNADLLQIASD 480
DB 421 LLANDIDNSNPVQAEQNLWHLVNYGSIIVANDPEANFDGVRVADVNNADLLQIASD 480
QY 481 YLKAHYGVDDKSEKNAINHLSILEANSNDPQNKOTKGAQLPIDNKLRLSLYALTRPLE 540
DB 481 YLKAHYGVDDKSEKNAINHLSILEANSNDPQNKOTKGAQLPIDNKLRLSLYALTRPLE 540
QY 541 KDAASNKNEIRSGLEPVITNSLNRSABGKNSRMANYIFIRAHDSVQTVIAKIIKAAQIN 600
DB 541 KDAASNKNEIRSGLEPVITNSLNRSABGKNSRMANYIFIRAHDSVQTVIAKIIKAAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYVTSQNTPTAVALMLSNKDSITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYVTSQNTPTAVALMLSNKDSITRLYYGDMYS 660
QY 661 DDQGYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DDQGYMATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDOGESEATKQGMVITNSNPSLKLNDKQIVNMGAAHKNQYRPLLLTTKQGL 780

Db 721 GANEATDQSEATKTOGMAVITSNPNLSKLNQNDKIVNVNMGAAHKNQYRPLLLTTKQGL 780
QY 781 TSYTSDAAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVPPVGASDNQDVRVAAS 840
Db 781 TSYTSDAAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVPPVGASDNQDVRVAAS 840
QY 841 NKANATGOVYESSALDSOLIYEGFSNFODFVTKDSYTNKKTIAQNVQLFKSWGVTSPFM 900
Db 841 NKANATGOVYESSALDSOLIYEGFSNFODFVTKDSYTNKKTIAQNVQLFKSWGVTSPFM 900
QY 901 APOYVSSDSGSLDJSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGIOVIAD 960
Db 901 APOYVSSDSGSLDJSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGIOVIAD 960
QY 961 WYPDQIYNLPKGEVWTATRVNDYGEYRKDSEIKNTLYAANTXSKNGKDYQAKYGGAPLSEL 1020
Db 961 WYPDQIYNLPKGEVWTATRVNDYGEYRKDSEIKNTLYAANTXSKNGKDYQAKYGGAPLSEL 1020
QY 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYPNGTNIILGRGVGYVLDKNASDKYFELAG 1080
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYPNGTNIILGRGVGYVLDKNASDKYFELAG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMYGL 1140
QY 1141 QQLNGEVQVFLSNGVOLRESFLENADGSKNYFGLHGNRYSGNYSPDNCKWRYFDASGV 1200
Db 1141 QQLNGEVQVFLSNGVOLRESFLENADGSKNYFGLHGNRYSGNYSPDNCKWRYFDASGV 1200
QY 1201 MAVGLKTINGNTQYFDDGYOVKGAITGSDGKKRYFDDGSGNMAVNRANDKNGDWYVL 1260
Db 1201 MAVGLKTINGNTQYFDDGYOVKGAITGSDGKKRYFDDGSGNMAVNRANDKNGDWYVL 1260
QY 1261 NSDGIALVGQTINGNTYTFQDGKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320
Db 1261 NSDGIALVGQTINGNTYTFQDGKQIKGKIITDNGKLYFLANSSELARNIPATDSQNN 1320
QY 1321 WYFVSGDVAVTGSQTIAGKLYFASDGKQVKGVSFVYNGKVHYHADSGELQVNRFEAD 1380
Db 1321 WYFVSGDVAVTGSQTIAGKLYFASDGKQVKGVSFVYNGKVHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYVLDNSGEALTGSORINDORVFFTRGKQVKGVDVAYDERRLLVYR 1430
Db 1381 KQGNWYVLDNSGEALTGSORINDORVFFTRGKQVKGVDVAYDERRLLVYR 1430

RESULT 4

US-09-740-274-2
Sequence 2, Application US/09740274
Patent No. US20020031826A1
GENERAL INFORMATION:
APPLICANT: Nichols, Scott E.
TITLE OF INVENTION: Glucan-containing Compositions and Paper
FILE REFERENCE: 0357CRD
CURRENT APPLICATION NUMBER: US/09/740, 274
CURRENT FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: 09/210,361
PRIOR FILING DATE: 1998-12-11
PRIOR APPLICATION NUMBER: 09/007,999
PRIOR FILING DATE: 1998-01-16
PRIOR APPLICATION NUMBER: 08/478,704
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: 09/009,620
PRIOR FILING DATE: 1998-01-20
PRIOR APPLICATION NUMBER: 08/485,243
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: 09/008,172
PRIOR FILING DATE: 1998-01-16
PRIOR APPLICATION NUMBER: 08/482,711
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
Query Match
Best Local Similarity 51.2%; Score 3817.5; DB 3; Length 1475;
Matches 762; Conservative 230; Mismatches 395; Indels 53; Gaps 22;
QY 1 METKERYKMHKVKHGVTVAVASGLITLTGTTTLLGSSVSAETBQOTSDKVVYTKSEDDKAA 60
Db 1 MDKKVRYKLRKVKRQVTVSVA VMTL-TTLGGGLVKADSNESKS-----QISNDSNFS 54
QY 61 SESSQTDAPKTKQAOTEQTOAQSOAN-VADTSTSTITKETPSONIITTOANSDDKTVTNTKS 119
Db 55 VVTANEESNVITEATSKQEAASSQTNHTVTTSSSSTSVVNPKEVSNPYTVGETASNGEK 114
QY 120 BEAQTSEERTKQSEEAQTTASSQALTQAKAEILTQRTAAQENKPNVDLAAIPNVKOIDG 179
Db 115 LQNTQITTV-DKTSEAAANNISKQT-TEADTVDDSNAA-----NLQILEKLPNVKEIDG 167
QY 180 KYIYIGSDGQPKKNFALTNNKVLVYDKNVTGALTDTSS-QYQFKQGLTKLNNDYTPHNQIV 238
Db 168 KYIYDNDNGKVRTNFTLIADGKILHFE-TGAYTDTSDTVNKDVTITRSNLYKKNQYV 226
QY 239 NFENTSLETIDNYVTADSWYRPKDILKNGKWTASSEDLRPLLMWSWPDKOTQATLNY 298
Db 227 DRSAQSFHFVDHYLTAESWYRPKYILKDGKWTQSTQTEKDFRPLMTWNPDPQSTQRYNY 286
QY 299 MNQOGLTGCTENTADSSOESLNLAQTVOVKLETKISOTQOTQWLRLDIINSFVKTOPNN 358
Db 287 MNAQ-LGINKITVDDTSNQLNIAAATQAKTEAKITTLKNTDMLRQTISAFVKQSAWN 345
QY 359 SOTESDTSAGEKHLQGGALLYNSDK-TAYANSYRILNRTPTTSQTK--PKYFEDNSS 415
Db 346 SDSEKPPD----DHLQNGAVLYDNEGKLTTPYANSYRILNRTPTNQTKKDPRTADNTI 401
QY 416 GGYDFLLANDIDNSPNVQAEQLNWLHLYMNVGSIIVANDPEANFGCVRVDAVDNVNADLL 475
Db 402 GGYDFLLANDVDNSPNVQAEQLNWLHFLMNFNGIYANDPDANFDSIRVDAVDNVADALL 461
QY 476 QIASDYLKAHYGVDSKSEKNAINHLSITLEAWSNDQYKNDTKGAQLPDKNKLRLSLYAL 535
Db 462 QIAGDYLKAAKGIHNDKANDHLSITLEAWSNDTPYLDHDDGDNWINDNKLRLSLFSL 521
QY 536 TRPLEKDSKNKNEIRSGLEFVITNSLNRSASGKNSERMANYIFIRAHDSVQTVTAKII 595
Db 522 AKPLAQ-----RSGMNPITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLIADI 573
QY 596 KQAINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPATYALMLSNKDSITRLYY 655
Db 574 KAEINPNVVGYSFTMEIEKKAPEIYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVY 633
QY 656 GDMVSDGQYMATKSPYVDAITLLKARIKYAAGQDMKIITYVEGDKSHMDWDYTCVLTS 715
Db 634 GDMFTDDGQYMAKHTINYEAETLLKARIKYVGGQAMRNQVGNSE-----IITIS 684
QY 716 VRYGTGANEATDQGEATKTOGMVITSNPNLSKLNQNDKIVNVNMGAAHKNQYRPLLLT 775
Db 685 VRYGKALKATDTGDRTTTSTSGVAIVEGNNPSRLKASDRVVVNMGAHKNQYRPLLLT 744
QY 776 TKDGLTSTYSDAAAKSLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVPPVGASDNQDV 835
Db 745 TDNGIKAVHSDQEAAGLVRYTNDRGELIPTAADIKGYANPQVSGYLGWVPVGA---LI 801
QY 836 RVAASNKANATGOV--YESSSALDSOLIYEGFSNFODFVTKDSYTNKKTIAQNVQLFKSW 893
Db 802 KMFALRLARPHQOMASVHQNAALDSKVMFEGSFNFOAFATKKEEYTNVIVAKNDVFAEW 861
QY 894 GYTSFEMAPQYVSSDSGSLDJSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKS 953
Db 862 GYTDFEMAPQYVSSDSGSLDJSIIQNGYAFEDRYDLGISKPNKYGTADDLVKAIKALHKS 921

QY 954 GIQVIADWVPDQIYNLPKGEVVTATRVNDYGEVRKOSSEIKNTLYAANTKSGNDYQAKYG 1013
DB 922 GIKVMADWVPDQMYAFPEKEVVTATRVNDYGEVRKOSSEIKNTLYAANTKSGNDYQAKYG 981
QY 1014 GAFLELAAKYPSPFNRTQISNGKKIDPSEKIATAWKAFYFNGTNILGRGVGYVLKDNASD 1073
DB 982 GAFLELAAKYPSPFNRTQISNGKKIDPSEKIATAWKAFYFNGTNILGRGVGYVLKDNASD 1041
QY 1074 KYFELKNGQ--TYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVODAKGNWYFDP 1131
DB 1042 TYFNISDNKEINFLPKTLNQDSQVGSYDGKGVYVYSTSGYQAKNTFISEG-DKWYFDP 1100
QY 1132 NNGHMVYGLQOLNGEVOYFISNGVQLRESFLENADGSKNYFGLHGRNYSNGYISFNDNSK 1191
DB 1101 NNGYMTGAQISNGVYVYFISNGVQLRESFLENADGSKNYFGLHGRNYSNGYISFNDNSK 1159
QY 1192 WRYFDASGMAVGLKTINGTQYFDQDGYQVKGAWITGSGDKKRYFDGSGMAVNRFPAN 1251
DB 1160 WRHFN-NGEMSVGLTVIDGQVYFDGQVYFDGQVYFDGQVYFDGQVYFDGQVYFDGQVY 1218
QY 1252 DKNGDWYILNSDIALVGVQTINGKTYVFGDCKQIKGKIITD-NGKLYFLANSCELAR 1310
DB 1219 NEEKWLILGEDAAVTGSGTNGQHLVFRANGVQVKGFEVTDHGRISYIDGNSGDQIR 1278
QY 1311 NIPATDSQNNWYFSGDGVAVTGSQTIAGKGLYFASDGQVKGSGFVTYN-GKVHYTHADS 1369
DB 1279 NRFRVNAQGWYFDDNNGYAVTCAITNGQLLYFRANGVQVKGFEVTDHGRISYIDGNS 1338
QY 1370 GELQVNRPEADKGNWYILDSNGEALTGSGRINDQRVFTREKQVKGVDVAYDERLLVY 1429
DB 1339 GQIIRNRFRVNAQGWYFDDNNGYAVTCAITNGQHLVFRANGVQVKGFEVTDHGRISY 1398

RESULT 5
US-10-383-930-34
; Sequence 34, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-34

Query Match 51.2%; Score 3817.5; DB 4; Length 1475;
Best Local Similarity 52.9%; Pred. No. 9.7e-190;
Matches 762; Conservative 230; Mismatches 395; Indels 53; Gaps 22;
QY 1 METKRYKMKYKQKHWTVAVASGLITLGTTLGSSVSAETBQOTSDKVVTOKSDDKAA 60
DB 1 MDKRYKRLKRYKRWTVSVASAVWL--TTLGGLVKADSNESK-----QISDNTS 54
QY 61 SSSQTDAPKTQAOEQTEQTAQSQAN-VADTSTSTKETPSPQNTTQANSDDKTNTKS 119
DB 55 VVTANESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKEVSNPVTGETAANGK 114
QY 120 EBAQTESEBTKQSEAOITASSQALTOAKAELTKORQTAQENKAPVDLAALPNVKQID 179
DB 115 LQNTQTTV-DKTSEAAANNSKQT-TEADTDVIDDSNAA-----NLQILEKLPNVKEIDG 167

QY 180 KYIYIGSDGQPKKFNALTVNNKVLVYFDKNTGALTDTST-QYQFKGLTKLNDYTPHNPQIV 238
DB 168 KYIYINDNNGKVRNFTLIADGKILHFE--TGAVTDSIDTVNKDIVTTRNSLYKYNQVY 226
QY 239 NFENTSLETIDNTVTADSWTRPKDILKNGKWTATASSEDRLPLMSWPKQTOIAYLNY 298
DB 227 DRSAQSFHEVDHYLTATSWTRPKYILKDGKWTQSTEKDFRPLMTWMPDQETQRYVNY 286
QY 299 MNQGLGTGNTYADTSOESLNLAACQTVQVKIETKISQTOOTOWLRDINSFVKTPNNW 358
DB 287 MNAQ-LGINKTYDDTNSQNLNAAATIOAKI EAKITLTKNTDMLROTISAFVQTOSAWN 345
QY 359 SOTESDTSAGEKHLGQALLYNSDK-TAYANSDYRLNRTPTSTQK--PKYFEDNSS 415
DB 346 SDEKFPD---DHLQNGAVLYDNEKLTPIYANSYRILNRTPTNQTKGKDPRTADNTI 401
QY 416 GGYDFILLANDIDNSNPVQAEQNLWHLNYGSIIVANDPEANPDGVVRVDAVNNVADLL 475
DB 402 GGYEFFLLANDVDSNPVQAEQNLWHLNFGNIYANDPDANFDSIRVDAVNNVADLL 461
QY 476 QIASDYLKAHYGVDKSEKNAIHLSTLEAWSNDNDPOYNKDTKGAQLPIDNKLRLSLAYAL 535
DB 462 QIAGDYLKAAKGIHKNDKKAANDHLSILEAWSNDNTPYLDHDDGDNMNMNDKRLSLFSL 521
QY 536 TRPLEKDASNKNEIRSGLEPVIITNSLNRSAGKNSERMANYIFIRAHDSVOTVIAKII 595
DB 522 AKPLNQ-----RSGMPLITNSLVNRTDDNAETAAPVPSYFIRAHDSVQDLADI 573
QY 596 KQAINPKTDLFTLDELKQAFKIYNEDMRQAKKYTQSGNIPTAYALMSNKSQITRLYY 655
DB 574 KAEINPVVGVSYFTBEIKKAFIYNKOLLATEKKTHTYNTALSYALLTNKSSVPVY 633
QY 656 GMYSDGQYMATKSPYDAIDTLKARIKYAAGGQDMKITVYEGDKSHMDWDTYGVLT 715
DB 634 GDMFTDGGYMAHTKINYEAIETLLKARIKYVSGGQMRNQVGNSE-----IITS 684
QY 716 VRYGTGANEATDQGESEATKTQGMVITSNNSPSLKNQNDKVIINMGAAHKNQYRPLLT 775
DB 685 VRYGKALKATDGTDRITTSVAVTEGNNPSLRLKASDRVVVNMGAHKNQYRPLLT 744
QY 776 TKDGLTSYTDAAAKSLYRKTNDKELVFDASDIQGLNPNQVSGYLAUVVPVQASNDQV 835
DB 745 TDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLGWVPVGA--LI 801
QY 836 RVAASNKANATQV--YESSALDSQLIYEGESNQDFVTKDSQDYNKKIAQNVQLPKSW 893
DB 802 KMFALLRAPHQOQMASVHQAALDSRVMFEGFSNFQAFATKBEYTNVVIKAVDXKFAEW 861
QY 894 GVTSEMAPQYVSSYEDSGSPLDSIIIONGYAFEDRYDLAMSKNNKYSQOQMINAVKALHS 953
DB 862 GVTDFEMAPQYVSSYEDSGSPLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKRAIKALHS 921
QY 954 GIQVIADWVPDQIYNLPKGEVVTATRVNDYGEVRKOSSEIKNTLYAANTKSGNDYQAKYG 1013
DB 922 GIKVMADWVPDQMYAFPEKEVVTATRVNDYGEVRKOSSEIKNTLYAANTKSGNDYQAKYG 981
QY 1014 GAFLELAAKYPSPFNRTQISNGKKIDPSEKIATAWKAFYFNGTNILGRGVGYVLKDNASD 1073
DB 982 GAFLELAAKYPSPFNRTQISNGKKIDPSEKIATAWKAFYFNGTNILGRGVGYVLKDNASD 1041
QY 1074 KYFELKNGQ--TYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVODAKGNWYFDP 1131
DB 1042 TYFNISDNKEINFLPKTLNQDSQVGSYDGKGVYVYSTSGYQAKNTFISEG-DKWYFDP 1100
QY 1132 NNGHMVYGLQOLNGEVOYFISNGVQLRESFLENADGSKNYFGLHGRNYSNGYISFNDNSK 1191
DB 1101 NNGYMTGAQISNGVYVYFISNGVQLRESFLENADGSKNYFGLHGRNYSNGYISFNDNSK 1159
QY 1192 WRYFDASGMAVGLKTINGTQYFDQDGYQVKGAWITGSGDKKRYFDGSGMAVNRFPAN 1251
DB 1160 WRHFN-NGEMSVGLTVIDGQVYFDGQVYFDGQVYFDGQVYFDGQVYFDGQVYFDGQVY 1218
QY 1252 DKNGDWYILNSDIALVGVQTINGKTYVFGDCKQIKGKIITD-NGKLYFLANSCELAR 1310

Db 1219 NEEGKWLVLGEDAAVTGSQTINGQHLVFRANGVQVKGFEVTDHGRISYYDGNQSDQIR 1278
QY 1311 NIFATDSQNNWYFSGDVAVTGSQTIAGKGLYFASDQKGVKGSFVTYN-GKVHYTHADS 1369
Db 1279 NREVRNAQGWYFEDNNGYAVTGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDGN 1338
QY 1370 GELQVNRFPADQGNWYFLDSNGEALTGSRINDQVFFTRREGKQVKGDAVDERLLVY 1429
Db 1339 GDQIRNRFVRNAQGWYFEDNNGYAVTGARTINGQHLVFRANGVQVKGFEVTDHGRISY 1398

RESULT 6

US-10-797-821-34
; Sequence 34, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-34

Query Match 51.2%; Score 3817.5; DB 5; Length 1475;
Best Local Similarity 52.9%; Pred. No. 9.7e-190;
Matches 762; Conservative 230; Mismatches 395; Indels 53; Gaps 22;

QY 1 METKRYKHKVKKHVTVAASGLITLGTTLGSSVSABTEQOTSDKVVTOKSEDDKAA 60
Db 1 MDKKVRYKURVKRWVTVSASAVMIL-TTISGGLVKKADSNESK-----QISDNTS 54
QY 61 SESSOTDAPKTKOAEQOTQAOQAN-VADTSTSIKTPSONITTOQANSDDKTVTNTKS 119
Db 55 VVTANEESNVITEATSKQEAASSQTNHTVTTSSTSVVNPKEVSNPVTGSETASNGEK 114
QY 120 BEAQTSEERTKQSEBAQTASSQALTOAKAELTKQRTQAAQENKNPVDLAALPNVQIDG 179
Db 115 LQNTQTIV-DKTSEAAANNISKQT-TEADTDVDDSNAA-----NLQILEKLPNVKEIDG 167
QY 180 KYYIGSDGQPKNFALTVNNKLVYDKNTGALTDTNS-QYQPKQGLTKLNDVTPHNQIV 238
Db 168 KYYIDNNGKVRNFTLIADGKILHFDE-TGAVTDTSIDTVNKKIDIVTTSNLYKKNQY 226
QY 239 NFENTSLETIDNVTADSWYRPKDIKNGKTTWTASSESDLRPLLMSPWDPKQTQIAYLNY 298
Db 227 DRSAQSFHVHYLTAESWYRPKYILKDGKWTQSTEDKFRPLLMTWPDQETQRYVNY 286
QY 299 MQOQGLGTGENTYADSSQSLNAAOTVQVKIETKISQTOQOTOWLRIIDINSFVKTPQNN 358
Db 287 MNAQ-LGINKTYDDTSNQLQNLAAATIQAKIEAKITTLKNTDNLRTQISAFVKTSAWN 345
QY 359 SOTESDTSAGEKHLQGGALLYSNDK-TAYANSDYRLNRTPTSTGK-PKYPEDNSG 415

RESULT 7

US-10-383-930-38

```
; Sequence 38, Application US/10383930 /
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Taubman, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
; US-10-383-930-38

Query Match      48.4%; Score 3615; DB 4; Length 1554;
Best Local Similarity 49.9%; Pred. No. 3.5e-179; Mismatches 418; Indels 112; Gaps 32;
Matches 745; Conservative 217;

QY      1  METKRYKHKVKKHVVTVAVAS-GLITL---GTTTLGSSVSAETEQOTSDKVVYTKQSED 56
DB      1  MEKKLHYKHLKVKHVVTVIAVSLVGLVAGTVAEDKQVANDTTAQATVGVDYDTCQ--- 57

QY      57  DKAAESSQTDAPKT-----KQAGTEQTOAQS-QANVADTSTSTIKETPTSONITTOANSDD 111
DB      58  DQATTNDANTNTDTPADQSANNTNQDQSGDSNNQDQAKQDT-----ANTDR 106

QY      112  KTVNTKSEAOSEBRTKQSEBAQTASSQALTQAKAELTKQRTAAQENKMPVDLAAI 171
DB      107  NQADNSQTNQNTDQATSPATDGTQVQRDAANVATA-ADQGGTAPSEQEKSAAL-SL 164

QY      172  PNVKQIDGKYIYIGSDGPKQKGFALTVNNKVLVYFDKNTGALTDTQSYQFKQGLTKLNNDY 231
DB      165  DNVLKIDGKYIYVQADSGYKKNFAITVNGQMLYFDSDTGALSTSTYSFSQGTTLNVDDP 224

QY      232  TPNQIVNFENTSLFTIDNYVTADSWYRPKDLKNGKWTWTSSESRLRPLMSWPPDKQT 291
DB      225  SSHNKAYDSTAKSFELVNGYLTANSWYRPAGILRNGQWTWASNENDLRPVLMSWPPDKQT 284

QY      292  QIAYLNMNQOGLGTGENTYADSSQESNLAAQTVQVKLETIKTSQTOQOWLRDIINSFV 351
DB      285  QVAYVNMNKYLSANETEVTNETSQVDLNKEAQSIQTKIEQKITSNDNSTOWLRTAMEAFV 344

QY      352  KQCPNMNSQTESDTSAGEKDHQGGALLYSNSDKTAYANSYRLNRTPTSTQTKPKYPE 411
DB      345  AAQPKWNMSTE---NPNKGDHLQGGALLYTNSDLTPWNSDYRLNRTPTQDGTCKYFT 401

QY      412  DNSSGGYDFLLANDDNSNPPVQAEQLNLHLYLMYVGSIVANDPEANFDGVRVDAVDNVN 471
DB      402  EGGEQREYEPFLNSNDVNSPPVQAEQLNLHLYLMWGDIVMGDKDANFDGVRVDAVDNVN 461

QY      472  ADLLQIADSLYKAHYGVDSKSNATNHLNLSLEAWSNDNDPQYNKDTKGAOLPIDNKLRLSL 531
DB      462  ADLLQVSYNFQNDYKVTVDSEANALAHISILEAWSLNDNQYNEDTNGTALSIDNSRLTS 521

QY      532  LYALTR-PLEK-DASNKNEIRSGLEPVITNSLN-NRSABGKNSERMANYIFIRAHDSVQ 588
DB      522  LAVLTQKQFORIDLNS-----LISESVNKEKANTAYGDTIPTYSFVRAHDSVQ 571

QY      589  TVIAKIIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYKTOSNIPTAYALMSNKK 648
DB      572  TVIAKIVKEKIDTNSDGYFTLLDQLKDAFKIYNEDMAKVNKYTYHYNIPAAVALLISNNE 631

QY      649  SITRLYLYGDMYSDGQYMAKSPYDAIDTLKARIKAAAGQDQMKITTVGDKSHMDWD 708
DB      632  SVPRVYVYGLYTDGQYMAKSPYDAIDATMLQGRYAYVSGGQSEBHVHKNQ----- 686
```

```
QY      709  YTCVLTSVRYGTGANBATD-QGSEATKTOGMNAVITSNPPLKLNQNDKVIWNMGAAHKQ 767
DB      687  ---ILSSVRYGQDLMSADDTQGTDLSTGLVTLVSNPNLDLG-GBSLTVNMGRAHANQ 742

QY      768  EYRPLILTYKDLSTSYTSDAAAKSLYRKTNKXELVFDASDIQGYLNPQVSGYLAVWVPV 827
DB      743  AYRPLILGTGQGVQSYLKSDST-NIVKYTDANGNLTFDTADDIRKYSTVDMSGYLAVWVPV 801

QY      828  GASDNQDVRVAASNKANATGOVYESSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQNV 887
DB      802  GAKDGDVRAADTNQKADGKSLKTSAAALDSQVIYEGFSNFQDPANNDADYTNKKIAENA 861

QY      888  QLFKSWGVTSEFEMAPQYVSEDSFLDSIIQNGYAFEDRYDLAWSKNNKYGSOQDMINAV 947
DB      862  DFFKKLGITSEFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAWSKNNKYGSKDDLAL 921

QY      948  KALHKSGIQVIADWVDPQIYNLPCKEVTATRVNDYGEYRKDSEIKNTLYAANTKSKNGD 1007
DB      922  KALHANGIQAIADWVDPQIYQLPGEVVTAKRTNSYGNPTFDAYINNALLYATNTKSSGSD 981

QY      1008  YQAKYGGAFSLAELAAKYPSEIENRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVGYL 1067
DB      982  YQAKYGGAFDELAKYDPMTFTVMISTGKPIDPSTKIQWEAKYFNGTNVZLGAGYVL 1041

QY      1068  KDNASDKYFELKXGNTYLPKQMT-NKEASTGFVNDGNGMTFFYSTGYQAKNSFVQDAKGN 1126
DB      1042  SDDATGKYPTVNEGDFLPASFTGDQNAKTGFFYDGTGMAYYSTSGNKA VNSFIYEG-GH 1100

QY      1127  WYFDNNGHMYGLQQL-NGEVOYFSLNSGVLRESFLENADGSKNTFGLGNRY-SNGYY 1184
DB      1101  YYYFDKGHMVTSYKAEDGNDYIFLPLNGIQMRDAIYODAQGSYYIYGRITGILYKGDNY 1160

QY      1185  SF-----DNDSKRWYFDASGVMAVGLKTINGNTQYFDQDQGVQVKGAWITGSDGKKRYFDD 1239
DB      1161  PFVDPNNAKTVFRIPDANNVPAIGYRNMYGQTYIFDENGFQAKGQLLTDKG-THYFDE 1219

QY      1240  GSGNMAVRPANDKNGDWYLLNSDGIALVGVQTINGKTYTFG-QDGKQIKGKIITD-NGK 1297
DB      1220  DNGAAMAKNFVN-VGDDWYMDGNGNAVKGQYVNNQILYFNPETGVQVKGQFITDAQGR 1278

QY      1298  LKYFLANSSELARNIPAT-----DSQNNWYFG-SDGVAVTGS 1334
DB      1279  TSYDANSALKSSGFFTPNGSDWYYAENGYVYKGFQVAENQDQWYFYDQTTGKQAKGA 1338

QY      1335  QTIAGKGLYFASD-GKQVKSFTV-YNGKVHYHADSGLQVNRREADDGNWYLLDSNG 1392
DB      1339  AKVDGRDLIFNPDPGQVQVKGDFATDESNTSFTHGNDGDKVGVGFFTTGNNWYADNNG 1398

QY      1393  EALTQSQR-----NDQRVFF-TREGKQVKGDVAYD 1422
DB      1399  NLVKGFEIDGKWHFDEVYTGQAKGAALVNGQQLYFDVDSGLQVKGDFVTD 1450

RESULT 8
US-10-797-821-38
; Sequence 38, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
```


Db 389 YSNNSKLTQANSNRYLRNTPNTQTKGKDPRTYADRTIGGYEFLLANDVDSNPVVOAE 448
Qy 437 QLANWLHYLMYNGSIVANDPEANFQGVVRVDAVNVNADLLQIADSYLKAHYGVDSKSEKNAI 496
Db 449 QLANWLHFLMFGNIYANDPDANFDSIRVDAVNDVADLLQIAGDYLAAGIHKNDKAAAN 508
Qy 497 NLSILEAWSDNDPOYNKDTKGAQLPDNKLRLSLLYALTRPLEKADSNKNEIRSGLEPV 556
Db 509 DHLISILEAWSYNDTPYLLHDDGDNNMINNDNRLLSLLYSLAKPLNQ-----RSGNPL 560
Qy 557 ITNSLNRSAGKNSERMANYIFIRAHDSSEVQTVIAKIIKAIKNPKDGLTFTLDELKQA 616
Db 561 ITNSLVNRTDNRATAEAPVSYSPFIRAHDSSEVQDLIRNIIRTEINPNVVGYSFTTEIKKA 620
Qy 617 FKINEDMROAKKYYTOSNIPAYALMSNKDSITRLYYGDMYSDDDQGYMATKSPYYDAI 676
Db 621 FEIYNKDLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGDMPFTDDGQYMAHKTINYEI 680
Qy 677 DTLKARIKAAAGQDMKITYVEGDKSHMDWDYTGVLTSVRYGTGANEATDQGEATKQ 736
Db 681 ETLKARIKVVSGQAMRNQOVGNSE-----IITSVRYGKGALKATDGTDRTRTS 731
Qy 737 GMVITSNPSLKLNDKVIYVNGAAHKNQYRPLLLTTKDGITSYTSDAAKSLYRKT 796
Db 732 GVAVIEGNPSLRKLKASDRVVVNGAAHKNQARPLLLTTDNGIKAYHSDQEAAGLVRYT 791
Qy 797 NDKGELVFDASDIQGYLNPQVSGYLAVWVPVGASDNDQVRVAASNKANATGVYVESSSAL 856
Db 792 NDRGELIFTAADIKGVANPQVSGYLVWVPVGAADQDVRVAASTPSTDGKSHVQNAAL 851
Qy 857 DSQILIEGFSNFQDFTVKDSYNTKKIAQNVQIPLKSGWVTSFEMAPQYVSESDGSLDSI 916
Db 852 DSRVMEFEGSNFQAFATKKEEYTVVIAKNVDKFAEWGVTDFEMAPQYVSSDTSGLDSV 911
Qy 917 IQNGYAFEDRYDIAMSNNKYGSOQDMINAVKALHKSIGIOVIADWVPDQIYNLPGKEVVT 976
Db 912 IQNGYAFTRDYDIGISKNPKYGTADLLVKAIKALHKSIGIKWADWVPDQIYNLPGKEVVT 971
Qy 977 ATRVNDYGEYRKDSSEIKNTLYAANTKNSGKDYOAKYGGAPLSLAAKYPSIFNRTOISNG 1036
Db 972 ATRVDKYGTPVAGSQIKNTLYVDGKSGSKDQKQYGGAPLELQAKYFELFARKQISTG 1031
Qy 1037 KKTDPSEKITAWAKYFNGNINILGRGVYVYKDNASDKYFELKGNQTYLPKQMTN----- 1091
Db 1032 VPMDPSVKIKQWSAKYFNGTNILGRGAGYVYKQATNTYFSLVSDNTFLPKSLVNPNGHT 1091
Qy 1092 KEASTGTVNDGNGWTFYSTGYOAKNSFVQDAGKNYFYDNGHMYVGLQOLNGEVQYFL 1151
Db 1092 SSVTVGLVFDGKGYVYSTSGNQAKNAFI-SLGNMYFYDNGMYVMTGAQSGINGANYFL 1150
Qy 1152 SNGVQLRESFLENADGSKNYFGLGNRYSGNYYSFNDNSKWRVYFDASGVMAVGLKTINGN 1211
Db 1151 SNGIQLRNAIYDNGKNVLSYNGDGRYENGYYLF--GQWRVYFQ-NGIMAVGLTRVHGA 1207
Qy 1212 TVQFDQDGYQVKAWITGSDGKRYFDDGSGNMAVNRFANDKNGDYYLNSDGIALLVGVQ 1271
Db 1208 VQYFDASGFQAKQOFTTADGKLYRFDSDSGNQISNRFRVNSKGEWFLFDHNGVAVTGTV 1267
Qy 1272 TINGKTYYPQDGKQIKGLIITD-NGKLYFLANSSELARNIPATDSQNNWYFYFGSDGVA 1330
Db 1268 TFGQRLYFKPNQVQAKGEFIRDANGLYRYPDPSNGNEVRNREVRNSKGEWFLFDHNGIA 1327
Qy 1331 VTGSQTIAG 1339
Db 1328 VTGARVNG 1336

RESULT 12

US-10-383-930-37

; Sequence 37, Application US/10383930

; Publication No. US20040127400A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
; US-10-383-930-37

Query Match 47.9%; Score 3575; DB 4; Length 1590;
Best Local Similarity 48.4%; Pred. No. 4,3e-177;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

Qy 1 METKRYKXHKVKKHVVAVASGLITLTGTTTIGSSVSAETEQTSDKVVTQKSEDDKAA 60
Db 1 MEKNVRFKHKVKKRWVTLVSASA--TMLASALGASVAS-----ADTFTAS 44
Qy 61 SESSQTDAPKTKQAQTEQTQAQSQANVADTSTSTIKETPSQNTTQANSDDKTVNTKSE 120
Db 45 DDSNQA---VVTGDTQTNNAQTDTSIATATS-EQSASTDAATDQASAEQTQGTAST 100
Qy 121 EAQTSERTKQSEAO---TTASSQALTOAKABELTKORQTAQAKENKPNVDLAAIPNVKQI 177
Db 101 D--TAAQTITNANEAKWPTENENQGFDEMLAEAKNVAT-AESDSTPSDLAKMSNVQV 157
Qy 178 DGKYYIGSDGQPKNFALTNNKVLVFDKNTGALTDTDSQYQFKQGLTKLND---YTPH 234
Db 158 DGKYYIYDQGNVKNFVSVGDKIYYFDE-TGAYKDTSKVDADKSSASVQNAITPAAN 216
Qy 235 NQIVNPENTSLIEDINVTADSWYRPKDLKNGKWTATASSESLRLPLLSWPDQTOIA 294
Db 217 NRAYSTSAKNFEAVDNYLTADSWYRPKSLKDGKWTESGKDDFRPLLMAMWEDTETKN 276
Qy 295 YLYNMNQGLGTCENTADSSQESLNLAAGTVQVQIETKISQTOQTOWLRDIINSFVKQ 354
Db 277 YVYMNKV-VGIDKTYTAETSQADLTAAAEVQARIEQKITSNNNTKWLREALSFAFKTQ 335
Qy 355 PNWNSQTESDTSAGEKDHLOGGALLYSN-SDKTAYANSDYRLNRPPTSGT-----KPKY 409
Db 336 PQMGSEKPYD---DHLQNGALLFDNQDTLTPDTQSNRYRLNRPPTNQTSGLDSRFTY 391
Qy 410 FEDNSGGYDFLLANDIDNSNPVQAEQLNWLHLNANYGSIIVANDPEANPDGVVRVDAVN 469
Db 392 NPNDPLGGYDFLLANDVDNSNPVQAEQLNWLHLNFGSIYANDADANFDSIRVDAVN 451
Qy 470 VNADLLQIASDYKLAHYGVDSKSEKNAIHLSTILEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529
Db 452 VDADLLQISSDYLLKAAAGIDKKNKANNHVSIVAEWSNDNTPYLHDDGDNLMNMNKFRL 511
Qy 530 SLLYALTRELEKXAKNKEIRSGLEPVTNSLNRSAGKNSERMANYIFIRAHDSSEVQT 589
Db 512 SMLWSLAKPLDK-----RSLGNPLIHSNLDREVDDREVETVPSYFARAHDSSEVQD 563
Qy 590 VTAKIKAKQINPKTDGLTFTLDELQAKFIYNEDMRQAKKYTOSNIPAYALMLSNKDS 649
Db 564 IIRDIKAEINPNSFGYSFTQBEIEQAFKIYNEDLKKTKKTHYVNPVLSYLLLTNKG 623
Qy 650 ITRLYYGDMSDDGYMATKSPYYDAITLLKARIKAAAGQDMKITYVEGDKSHMDWDY 709
Db 624 IPRVYVYGDVMTDDGQYMAKNTVNYDAIESLLKARMKYVSGGQAMQ-NYQIGNCE----- 676
Qy 710 TGVLTSVRYGTGANEATDQGEATKQGMVAVITSNPSLKLNDKVIYVNGAAHKNQY 769
Db 677 --ILTSVRYGKALKQSKGDATRTTSGVGVVNGNQPNFSLD-GKVVALNMGAHANQY 733

```
QY 770 RPLLLTTKDLGTSYSDAAK--SLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVVPV 827
Db 734 RALMVSTKDGVTATYATDADASKAGLVKRTDENGYLFLNDDLKGVANPQVSGFLQVWVPV 793
QY 828 GASDNQDVRVAASNKANATQGVYSESSALDSQLIYEGFNGFQDFVTKDSQDYNTKKIAQNV 887
Db 794 GAADDQDIRVAASDRTASTDGKSLHQDAAMDSRVMPFEGFNFQSFATKESEYTNVVIANNV 853
QY 888 QLFKSGVTSFEMAPQVYSEDSGLDSIQNGYAFEDRYDLAMSKNNKYSGOODMINAV 947
Db 854 DKFVSGITDFEMAPQVYSSDQFLDSVQNGYAFTRDYDLGMSKANKYGTADQLVKAI 913
QY 948 KALHKSGIOVIADWVDPQIYNLPCKEAVTATRVNDYGEYRKDSEIKNTLYAANTKSKNGD 1007
Db 914 KALHAKGLKVMADWVPDQWYTPFKQBEVTVTRDKFKGPIAGSQIHNSLYVTDITKSSGDD 973
QY 1008 YQAKYGGAFLELAELAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNNILGRGVGYVL 1067
Db 974 YQAKYGGAFLELKEKYPFLTKQISTQGAIDPSVKIKQWSAKYFNGSNILGRGADYVL 1033
QY 1068 KXNASDKYFELKGNQYLPKQMTNKEASTGFVNDGNGMTFYST--SGYQAKNSFVQDAKN 1126
Db 1034 SDQVSNKYFNVASDTLFLPSSLGKVVESGIRYDGGKGIYNSSATGDQVKASPIEA-GN 1092
QY 1127 WYFDPNNGHWYGLQOLINGEVQYFLSNGVQLRESFLENADGSKNYFCHLGNRYSGYISF 1186
Db 1093 LXYFPGKDGVMYVTAQTINGANYFPLENGTALRNTIYTDAGNSHYANDGKRYENGYYQQF 1152
QY 1187 DNDKRWYFDASGVMAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAV 1246
Db 1153 GND--WRYF-KDGNMAVGLTVDGNVQYFDKQGVQAKDLIVTRDGKRYFDOHNGNAAT 1209
QY 1247 NRPANDKNGDWYILNSDGLALGVQTINGKTYFYFGDGKQIKGIIT-DNGKLKYFLANS 1305
Db 1210 NTFADKTHWYILGKQGVAVTCAQTQVKGKLYFEANGQVKGDFVTSDEGKLYFYDVS 1269
QY 1306 GELARNIFATDSQNNHYFGSDGVATGSGQTITAGKLYPASDGKQVKGFSVT-YNGKVHY 1364
Db 1270 GDMWTDFTIEDKAGNWFYLGKGAAVTGAQTIRGQKLYFKANGQQVKGDIVRGTDGKIRY 1329
QY 1365 YHADSLEQVNRPEADKQ----- 1383
Db 1330 YDAKSGEQVFNKTVKAADGKTYVIGNDGVAVDPSVVKGQTFKDAQSALRPFYNLKGQIVTG 1389
QY 1384 -----NWYLDNSGALTSQRIHQVDFVTFREGKQVKGDV 1419
Db 1390 SGWYETANHDWVYIQS-GRALTGEQTINGQHLFYKEDGHQVKGQL 1433
```

RESULT 13

```
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match          47.9%; Score 3575; DB 5; Length 1590;
Best Local Similarity 48.4%; Pred. No. 4.3e-177;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

QY 1 METKERYKMHKKYKHWTVAVASGLITLTGTTTGLSSVSAAETEQOTSDKVVTKSEDDKAA 60
Db 1 MEKNVRFMKHKYKHWTVLSVASA--TMLASALGASVAS-----ADTDITAS 44
QY 61 SESSQTDAPKTKQAQTEQTAQSQANVADTSTSIKTETPSQNTITQANSDDKTVTNTKSE 120
Db 45 DDSNQA---VVTGDDQTTNNQATDQTSIAATATS-EQSASTDAATDQASAAEQTQGTAST 100
QY 121 EAQTSEERTKQSEBAQ----TTASSQALTOAKAELTKQRTAAQENKNPVDLAAIPNVKQI 177
Db 101 D--TAQTNTTNAEAKWPTENENOGFTDEMLAEAKNVAT-AESDSIPSDLAKMSNVKQV 157
QY 178 DGKYYIYIGSDGQPKNFALTNNKVLYPDKNTGALTDTTSOYQFKOGLTKLNND---YTPH 234
Db 158 DGKYYIYDQGNVKNFVSGVKIYYFDE-TGAYKDTSKVDADKSSVASQVATIPAA 216
QY 235 NQIVNFENTSLETIDNYVTADSWYRPRKOILKNGKTWTASSESDLRPLLSMWPDKQOTJA 294
Db 217 NRAYSTSAKNFEAVDNYLTADSWYRPKSILKDGKTWETSGKDDFRPLLMAMWPDTTETKRN 276
QY 295 YLYNNVQOGLGTGENVTADSSQESNLAAQTVOVKIETKISQTOQTOWLRDILNSFVKQ 354
Db 277 YVYNNKVV-VGIDKTYTAETSAQLTAAAEVLAELVQAREIKTSENNTKWLREAIASFVKQ 335
QY 355 PNWNSQTESDTSAGEKHLQGGALLYSN--SDKTAVANSDYRLNLRTPTSQTG---KPKY 409
Db 336 POWNGESEKPYD-----DHLQNGALLFDNOTDITPDQTSNYRLNLRTPTNQOTGLSDRFFY 391
QY 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNWLHLYLMNYGSI-VANDPEANFDGVRVDAVDN 469
Db 392 NFENDPLGGYDFLLANDVDNSNPVQAEQLNWLHLYLNFGSIVANDADANFDSIRVDAVDN 451
QY 470 VNADLLQIASDYLYKAHYGVDKSEKAINHLSILEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529
Db 452 VDADLLQIASSDYLYKAAYGIDKNNKANNHVSIVEAWSNDNTPYLDHDDGDNLMNMKNFRL 511
QY 530 SLLYALTRPLEKDAASKNKEIRSGLEPVI TNSLNRSAGKNGERMANYIPIRAHDSVOT 589
Db 512 SMLWSLAKPLDK-----RSLGPLIHNSLVDREVDDREVETVPSYSFARAHDSVQD 563
QY 590 VIAKIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKCYTQSNIPTAYALMLSNKDS 649
Db 564 IIRDIKAEINPNSFGYSFTQEEIEQAFKIYNEDLKTKDKKYTHYVPLSYTLTLTNKGS 623
QY 650 ITRLYYGDMSDQGYMATKSPYYDAIDTLKAKIKYAAGGQDMKITYVEGDKSHMDWY 709
Db 624 IPRVYVYGDFTDGGYMANKTNYDAIESLLKARMKYVSGGQAMQ-NYQINGE----- 676
QY 710 TGVLTSVRYGTGANEATDQSEATKTQGMNAVITSSNPNLSKLQNQNDKQVIVNMGAAHNQY 769
Db 677 --ILTSVRYGKALKQSDKGDATTRTSGVGVVGNQPNFSLD-GKVVALNMGAAHANQY 733
QY 770 RPLLLTTKDLGTSYTSDDAAK--SLYRKTNDKGLVFDASDIQGYLNPQVSGYLAVVVPV 827
Db 734 RALMVSTKDGVTATYATDADASKAGLVKRTDENGYLFLNDDLKGVANPQVSGFLQVWVPV 793
QY 828 GASDNQDVRVAASNKANATQGVYSESSALDSQLIYEGFNGFQDFVTKDSQDYNTKKIAQNV 887
Db 794 GAADDQDIRVAASDRTASTDGKSLHQDAAMDSRVMPFEGFNFQSFATKESEYTNVVIANNV 853
QY 888 QLFKSGVTSFEMAPQVYSEDSGLDSIQNGYAFEDRYDLAMSKNNKYSGOODMINAV 947
Db 914 KALHAKGLKVMADWVPDQWYTPFKQBEVTVTRDKFKGPIAGSQIHNSLYVTDITKSSGDD 973
```

Db 854 DKFVSGITDFEMAPQVSVSTGQFLDSVIQNGYAFTRDRLDGLMSKANKYGTADQLVKAI 913
Qy 948 KALHSGIQVIADVPDQIYNLPKGEVVTATRVNDYGEYRKDSBKNTLYAANTKSGKD 1007
Db 914 KALHAKGLKWDVVPDQMTFFKQEVVTVTRTDKFKPTAGSINSHLYVTDTKSSGDD 973
Qy 1008 YQAKYGAFIUSLAELAKPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNILRGVGYVL 1067
Db 974 YQAKYGAFIDELKEKPELFTKQISTGQAIDPSVKIKQWSAKYFNGSNILGRGADYVL 1033
Qy 1068 KDNASDKYFELKGNQTYLPKQMTNKKEASTGFVNDGNGMTFYST-SGYQAKNSFVODAKGN 1126
Db 1034 SDQVSNKYFNVAOTLFLPSLLGKVVESGIRVDGKGYIYNSSATGQVQKASFI TEA-GN 1092
Qy 1127 WYFDDNNGHMYGLQQLNGEVQVFLSNGVOLRESFLENADGSKNYFCHLGNRYSNGYYSF 1186
Db 1093 LYVFGKGYMVTGAQTNGANYFFLENGTALRNTIYTDAGNSHYANDGKRYENGVOQF 1152
Qy 1187 DNDKWRYPDASGVNAVGLKTINGTQYFPDQGVQVKGAWITGSDGKKRYFDDGSGNMAV 1246
Db 1153 GND--WRYF-KDGNMAVGLTVDGNVQYFDKQGVQAKDKIIVTRDGKRVYFDQHGNAAT 1209
Qy 1247 NRPAKNGKMGVYLNDSGIALGVQTINGKTYTFYFGQDGKQIKGLIIT-DNGKLYFLANS 1305
Db 1210 NTFIADKTGHYYLGGKDGVAVTGAQTVGKQKLYFEANGQQVKGDFVTSDEGLYFYDVS 1269
Qy 1306 GELARINAFATDSQNNWYFSGDVAVTGSQTIAGKKLYFASDGKQVKGSGFVT-YNGKVHY 1364
Db 1270 GDMWTFIEDKAGNFWYLGKDGAAVTGAQTIIRGQKLYFRANGQQVKGDIUVKGTGDKIRY 1329
Qy 1365 YHADSGLQVNRFEADKDG----- 1383
Db 1330 YDAKSGEQVFNKTKAADGTYTIGNDGVAVPSVVKGQTFKDSAGALRFPYNLKGQLVTG 1389
Qy 1384 -----NWYLDNSGEALTSQRIQNDQVFTFREGKQVKGDV 1419
Db 1390 SGWYETANHDWVYIQS--GKALTGEQTINGQHLFYKEDGHQVKGQL 1433

RESULT 14
US-10-383-930-40
; Sequence 40, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Taubman, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-383-930-40

Query Match 46.9%; Score 3502.5; DB 4; Length 1518;

Best Local Similarity 48.8%; Pred. No. 2.4e-173;
Matches 724; Conservative 247; Mismatches 434; Indels 101; Gaps 32;

Qy 1 METKRYKHKHKKHVVTVAVAS-GLITL--GTTILGSSVSAETEQTGDKVVTQKSEDD 57
Db 1 MENKIHKLHKVKKQVWTVAVASVALATVLGLSVTSSVSAD---ETQDKVTQ--SNSG 56
Qy 58 KAASESSQTDAPKTKQAQTEQTOAQ-----SQANVADTSTSTKETPSSQNTITQANSDDK 112
Db 57 TTASLTVTSPEATKADKRTNTEADVLTPAKETNAVETATTTNTQATAEAAATT-ATTADV 115

Qy 113 TVTNTKSEEAQTSEE----RTKQSEBAQTASSQAL--TOAKAELTKQRTAAQENKNPVD 167
Db 116 AVAAPNKEAVTVDAPAVTTEKAEQOPATVKAENVNTEVA-----PRAALKDSEVEA 169
Qy 168 LAAPNVKQIDGKYIYIGSDGQPKKFPALFVNNKVLVFDKNTGALTDTSOYQFKQGLTKL 227
Db 170 ALSLKNIKIDGKYIYVWEDSGHKNFAITVNGQLLYFGKD-GALTSSSTVSFTPGTTNI 228
Qy 228 NNDYTHNQIWNENTSLFTIDNYVTADSWYRPDKILKNGKWTWTASSESLRPLLMASWPP 287
Db 229 VDFGSINNRAYDSSEASFELIDGLTADSWYRPASIKDGVTWQASTAEDFRPLLMASWPP 288
Qy 288 DKQTOIAYLNNMQOGLGTGENTYADSSQESLAAQTVQVKLETKISQTOQWMLDII 347
Db 289 NVDTQVNYLNNYSKV-FNLDAKYSSTDKQETLVAAKDIQIKIEQKIQAESQWMLRETI 347
Qy 348 NSFVKTQPNWNSQTESDTSAGEXDLQGGALLYSNSDKTAYANSDYRLNRTTTSQTKG- 406
Db 348 SAFVKTPQWNKETENYKSGGSDHLQGGALLVYVDSRTPWANSDYRLNRTATNQGTI 407
Qy 407 -----PKYPEDNSSGGYDFLLANDIENSPPVQAEQNLWLHYLMNYSIIVANDEANPDG 461
Db 408 DKSILDEQSDPNHMGGFDFLLANDVLSNPNVQAEQNLQIHYLMNYSIIVMGDKDANFDG 467
Qy 462 VRVDADVNNADLIQIASDYLKAHYGVDKSEKNAIHLSTLEAWSNDNDPOYNKDTKGAOL 521
Db 468 IRVDADVNDADMLQYTNFYREYGVGNKSEANALAHISVLEAWSNDNDHNDKTDGAAL 527
Qy 522 PIDNKLRLSLYALTRPLEKASNKNEIRSGLEPVIITNSLN----- 562
Db 528 AMENKQRLALLFSLAKPIK-----ERTPAVSPLYNNPTNTTQORDEKTDWINKGSKAY 580
Qy 563 -----NRSAEGKNSE-----MANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFTFLDE 612
Db 581 NEDGTIVKQSTIGKYNEKYGDSAGNYVFIRAHDNNVQDIIAEIIEKKEINPKSDGFTIDAE 640
Qy 613 LKQAPKIYNEDMRQAKKKTOSNIPTAYALMSLNKDSITRLYYGDMYSDDGQYMATKSPY 672
Db 641 MKQAFIYNKDMSSDKKTLNIPAYAVMLQNMETITRYYGDDLYTDDGHGWETKSPY 700
Qy 673 YDAIDTLKARIKYAAGQGDQMKITVV--EGDKSHMD---WDYTGVLTSVRYGTGANEATD 727
Db 701 YDTIVNLMKSRIKYVSGGQQRSLPTDQKMDNSDVELYRTNEVTVSYRGKDIWNTAND 760
Qy 728 -QGSEATKTQCHAVITSNRPSLKLNDQKVIYVMGAAHKNQYRPLLLTKOGLTSYTS- 785
Db 761 TEGSKYSRTSGQVTLVANNPKLNDQSAKLANVEMGKIHANQYKRYALIVGTADGINKFTSD 820
Qy 786 -DAAAASLVRKTNDRKELVFDASDIQGLNLPQVSGVLAVWPVPGASDNQDVRVAASNKAN 844
Db 821 ADAIAAGYVKTETDSNGVLTFGANDIKGYETFDMSGFVAVWPVPGASDNQDIRVAPSTEAK 880
Qy 845 ATGQV--YESSALDSOLIYEGFSNFODFVTKDSD---YTNKIIAQNVLFKSWGVTSEFM 900
Db 881 KEGELTLKATEAYDSOLIYEGFSNFQ--IPDGSDPSTVYTNKIAENVDFLFKSWGVTSEFM 939
Qy 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHKSIGIQVIAD 960
Db 940 APQFVSADDTGDTFLDSVIQNGYAFADRYDLAMSKNNKYGSKEIDLKALHAKAGIQAIAD 999
Qy 961 WYPDOIYNLPKGEVVTATRVNDYGEYRKDSBKNTLYAANTKNGKDYQAKYGGAFSEL 1020
Db 1000 WYPDOIYQLPKGEVVTATRVNDYGEYRKDSBKNTLYAANTKNGKDYQAKYGGAFSEL 1059
Qy 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVYVVKDNASDKYFELKG 1080
Db 1060 KAKYEMFKVNMISTGKPIDDSVKLKQWKAEIFNGTNVLERGVGYVLSBETGKYFTVTK 1119
Qy 1081 NQTYLIPKQWNTKE-ASTGFVNDGNGMTFYSTSGYQAKNSFVQDQKGNWYTFDNNHMGVYG 1139
Db 1120 EGNFTPLQLTGKEKVTGFSDDGKITYFTGSGTQAKSAFV-TFNGNTYTFYDFARGHMTVN 1178

Db		1239	SKVVKERYFTNEGVMAGVTTVLDGFTQYFGEDGFQAOKLUVT-FKGKTYTFDHAHTNGXIK	1297
Qy		1247	NRFANDKNGDWYYILNSDGIALVGQTITNGKTYFFGDGKOIGKIITD-NGKLKYFLANS	1305
Db		1298	DTRN-INGKWYYFDANGVAATGAQVINGOKLYFNEDGSQVKGGVKNADGTYSYKEGF	1356
Qy		1306	GELARNIFATDSQNMYYPFSGDVAVTGSTIAGKKLYFPASDCQKVKGSPV-TYNQKVHY	1364
Db		1357	GELVTNEFFTQGNVWYIAGANKTIVGAQVINGOHLFYFNADGSQVKGGVKNADGTYSK	1416
Qy		1365	YHADSEGLQVRFEADKGNWYILDNSGEALTGTSORINDORVFPTREGKQVGD-VAYDE	1423
Db		1417	YNASTGERLTNEFFTQDNWYIIGANGKSVTGEVKIGDDTYFFAKDGQVKVQGTVSAGN	1476
Qy		1424	RRLLVY	1429
Db		1477	GRISY	1482

Search completed: February 11, 2006, 20:56:52
Job time : 134.994 secs

100

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 8.13754 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-36
Perfect score: 7462
Sequence: 1 METKRRYKHKVKKHWTVA.....EGQVKGVXDYDRLLYR 1430

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New.*
1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	451	6.0	2710	7 US-11-051-453-41	Sequence 41, Appl
2	393	5.3	2367	7 US-11-051-453-42	Sequence 42, Appl
3	370	5.0	726	7 US-11-052-554A-217	Sequence 217, App
4	296.5	4.0	690	7 US-11-052-554A-232	Sequence 232, App
5	280	3.8	1992	7 US-11-013-759-3	Sequence 3, Appl
6	280	3.8	1992	7 US-11-013-759-13	Sequence 13, Appl
7	280	3.8	2047	7 US-11-013-759-4	Sequence 4, Appl
8	280	3.8	2047	7 US-11-013-759-7	Sequence 7, Appl
9	275.5	3.7	1647	7 US-11-052-554A-260	Sequence 260, App
10	275	3.7	2314	7 US-11-013-759-11	Sequence 11, Appl
11	270	3.6	701	7 US-11-052-554A-231	Sequence 231, App
12	267.5	3.6	396	7 US-11-022-562-228	Sequence 228, App
13	266	3.6	5024	6 US-10-793-626-296A	Sequence 296A, App
14	265	3.6	619	7 US-11-052-554A-229	Sequence 229, App
15	265	3.6	658	6 US-10-873-528-17	Sequence 17, Appl
16	265	3.6	677	6 US-10-873-528-155	Sequence 155, App
17	265	3.6	1155	6 US-10-793-626-1780	Sequence 1780, App
18	257.5	3.5	627	6 US-10-873-528-191	Sequence 191, App
19	253.5	3.4	1107	6 US-10-485-517-145	Sequence 145, App
20	245	3.3	2053	7 US-11-013-759-9	Sequence 9, Appl
21	238	3.2	744	6 US-10-873-528-184	Sequence 184, App
22	235	3.1	2399	7 US-11-052-554A-92	Sequence 92, Appl
23	232	3.1	338	7 US-11-052-554A-228	Sequence 228, App
24	227	3.0	332	6 US-10-873-528-190	Sequence 190, App
25	227	3.0	5291	7 US-11-052-554A-281	Sequence 281, App

26	226.5	3.0	1263	6 US-10-485-517-127	Sequence 127, App
27	226.5	3.0	1448	6 US-10-485-517-212	Sequence 212, App
28	220.5	3.0	1290	6 US-10-485-517-141	Sequence 141, App
29	219.5	2.9	1562	7 US-11-052-554A-211	Sequence 211, App
30	219.5	2.9	3194	7 US-11-052-554A-90	Sequence 90, Appl
31	213.5	2.9	1094	6 US-10-517-939-144	Sequence 144, App
32	213.5	2.9	1095	6 US-10-793-626-3154	Sequence 3154, App
33	212.5	2.8	514	7 US-11-102-188-4	Sequence 4, Appl
34	212.5	2.8	693	6 US-10-873-528-185	Sequence 185, App
35	210.5	2.8	1579	7 US-11-052-554A-9	Sequence 9, Appl
36	209.5	2.8	1234	6 US-10-517-939-306	Sequence 306, App
37	208	2.8	512	7 US-11-102-188-5	Sequence 5, Appl
38	208	2.8	2871	7 US-11-124-367A-264	Sequence 264, App
39	206.5	2.8	1889	7 US-11-102-476-46	Sequence 46, Appl
40	206	2.8	1268	7 US-11-052-554A-1	Sequence 1, Appl
41	205.5	2.8	542	7 US-11-102-188-9	Sequence 9, Appl
42	205.5	2.8	1279	6 US-10-793-626-3188	Sequence 3188, App
43	205	2.7	484	7 US-11-102-188-16	Sequence 16, Appl
44	205	2.7	485	6 US-10-630-203-12	Sequence 12, Appl
45	205	2.7	485	7 US-11-113-775A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-PEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: Patentin Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 6.0%; Score 451; DB 7; Length 2710;
Best Local Similarity 20.0%; Pred. No. 9.7e-16;
Matches 330; Conservative 251; Mismatches 594; Indels 478; Gaps 77;
QY 86 NVADTSTITKTPSONITTOANSDDKTNTKSEEAQTSEERTKQSEA-OQTASSOAL 144
Db 720 SIMDKITSTLPDVKNKSITIGANQYEVRIINSECKELLAHSGKWINKEAINDLSKEY 779
QY 145 -----TQAKAELTKQRQTAAQENKNPVDLAAIPNVKQIDGKYIYIGSDGPKKNFALT 197
Db 780 IFFDSIDNKLKA-----KSKNIPGLASISE---DIKTLILDASVSPTKFIIL- 823
QY 198 VNNKVLVFDKNTGALTDTTSQYQKGLTKLNNDYTPHNOIVNFENFSLFTIDNYV----- 252
Db 824 -NNLKNIESSIG-----DYIYEKLEPVKNIHNSIDLDIDFNIL 863

Qy	253	---	TADSWYRPK-----	DILKNGKXTWTASSSDURLPPLMSWWPKQQTQIAY	295	
Db	864	LENYSDELYELKLNLDKYLISFEDISKNNSTYS-----	VRF	902		
Qy	296	LNYNQOGLGTGENYATYADSSQESLNLAQTVQVKIET-----	KISQQTQ	338		
Db	903	INKSN-----	GESVYETEKEIFSKYSEHITKEISTIKNSIITDVNGNLLDNQIDHTS	956		
Qy	339	QTOWLRIINSFVKTQPNWNSQTE--SDTSAGEKDHLOGGALLYSNSDKTAY-----	388			
Db	957	QVNTLNAAF--FIQSLIDYSSNKDVNDLSTSVKVLQY--AQLFSTGLNTIYDSIQLVNL	1012			
Qy	389	ANSDYRLNKTPTSQTGPKPKYFEDNSSGGYDFLLANDINSNPVQAE-----	QLN	439		
Db	1013	ISNAVNDTINVLPITTEGIPITVSTILGDCILNGAAIKELLDEHDEHPKKELEKAEKVGVLAIN	1072			
Qy	440	W-LHYLMNYSIVANDPEANPDGVRVDAVDNWNADLLOIASDYLKAHGVGDK--SEKVAI	496			
Db	1073	MSLSIAATVASIVGIGAEVT---IFLLPIAGISAGIPSLVNNELILH---DKATSVVNYF	1126			
Qy	497	NHLSILEAWSNDPDQYNKDTKGAQLPDNDKLRSLSLLYALTRPLEKOAASNEIRSGLEPV	556			
Db	1127	NHLSSEKYG--PLKTEDDK-ILVPIDDLV-----	TSEIDFNNSIKLG----	1167		
Qy	557	ITNSLNRSAGKNSERMANYIFIRAHDSVQTVIAKI--TKAQINPKTDGLTTLDELKQ	615			
Db	1168	--TCNILAMEGGSGHTVTGIDNHFFSSPSISHSHPSLISYSAIGIETENLDFS---KK	1220			
Qy	616	AFKIYNEDMRQAKKYYTQSNIPYAYALMSLNK---DSITRLY-----	YCDMYSDDGQYMAT	668		
Db	1221	IMMLPNAPSRVF--WNETGAVPLGSLRSLENDGTRLDLSIRDLYPGKFYMRFYAFDFVAIT	1278			
Qy	669	KSPYDAIDTLKARIKAYAGGQDMKITVEGDKSHMDVYTG-----	VLTSVRXYGTG	721		
Db	1279	LKPYE--DTNIKILKQDTRNFIMPITITNEIRKLSYSPDGAGGTYSILLSSYPISYN	1336			
Qy	722	ANEATDOGSEATKTQGMVITSNNPSLKLQNQDKVINYMGAHKQNYERPELLTTKDGLT	781			
Db	1337	INLSKDDLWIFNIDNEVREISIENTGIKKGKLIKDVLSKIDINKNK-----	LIIGNQTI	1390		
Qy	782	SYTSDAAAKSLY-----	RKTNDKGBELVFADSDIOQYLNPQVSGYLAVWVPVGAEDNQDVRV	837		
Db	1391	DFSGDINDKDRYIFELTCELDQKSLIIEINLVAKSYSLLSG-----	DKNY	1436		
Qy	838	AASKANATGOVYESSALDSQLI---YEGFSNFQDE--VTKQSDYT--	NKIAQNVQL	889		
Db	1437	LISLNSNTIEKI--NTIGLDSKNAYNTDESNNKYFGALISKYSQKSIHYHKQSKNILE	1494			
Qy	890	FKSGVTSFEMAPQVYSSSEDGSFL-DSI--IQNGYAFEDRYDLA-----	MSKN	934		
Db	1495	FYNDSTLEFN--SKDFIAEDINVMKDDINTITGKYVYDNDNTRKSIDFSISLVSKNVQKVN	1553			
Qy	935	-----	NKYGQQDMINAVKALHSG-----	IOVIADW-----	VPDQIYNLPGK	972
Db	1554	GLYLNESVSSYLDPFVKNSDGHHTNSFNMFLFDNISFWKLFGPENTFVIDKYFTLVGK	1613			
Qy	973	-----	EWYATRNV---DYGEYRKDSIEKNTLYAANTKS-----	NGKD-----	1007	
Db	1614	TNLGVYFEDCNKNNDIYFGEW--KTSKSKSTIFSGNGRNVVVEPIYNPTDGETISTSLD	1672			
Qy	1008	--YQAKYG-----	GAFLSLAAPKPSIFNRTQISNGKK	1038		
Db	1673	PSYEPFYGDYRINKVLIAPDLYTSLININTNYSNEYYPEIIVLNPETHKKVIN---	1729			
Qy	1039	IDPSEKITAWK-----	AKYFNGTN-----	ILGRG	1062	
Db	1730	LDSSSFYKWTSEGSDFILVRYLESNNKILQIRIKGILSNTQSFNKMSIDFDKDKKLS	1789			
Qy	1063	VGYVL-----	KDNASDK---YELKGNQTYLFPQMTNKEASTGCFVNDGNGMTFYS---	1109		
Db	1790	LGYTMSNPKSPNSENEDRDLHGFKPIDNKTYTYDE--DSKLVKGLININNSUFYFDPIE	1847			
Qy	1110	---TSGYQAKNSFVQADAKGNWYFED--NNGHMVYGLQOLNGEVOYFSLNGVOLRESFLENA	1165			

```

Db      1848  FNLVTGWQTIN-----GKYYFDINTGAALTSYKLIINGKHFFYFNDGV-MQLGVFKGP 1899
Qy      1166  DGSKNYFCHLGNRYN-----GYTSFNDISK-----YFDA 1197
Db      1900  DGEF-YFAPANTQNNIEGQAIYVQSKFLTLNGKKYYFDDNNSKAVTGWRIINNEKYYFNP 1958
Qy      1198  -SGMAVGLKTINGNTQYFDODGYQVGAWITSDGKKRYFDDGSGNMVNRFANDKNGD 1256
Db      1959  NNAIAAAGLVQIDNNKYYFNPDTAIIISKGWQT-VNGSRYYFDTDTA-IAFNGY-KTIDGK 2015
Qy      1257  WYILNSDGIALVGV-----QTINGKTYFFGQDGGK 1285
Db      2016  HFVDSDCVKVIGVFTSNGFEYFAPANTYNNIEGQAIYVQSKFLTLNGKKYYFDDNNSK 2075
Qy      1286  QIKGIITDNGKUKUYFLANGELARNIFATDSQNNWYFGSD-GVAVTSGQTAGKKLYF 1344
Db      2076  AVTGWOTIDSKKY-YFNTNTAEAAATGWOTIDGKK--YFNTNTAEAAATGWOTIDGKKYYF 2132
Qy      1345  ASDGQKVGSFVYNGKHVHYHADSGELQV-----NRPE-----ADCK----- 1382
Db      2133  NTNTASTGYTTINGK-HFYFTDGIQVLGVFGNGFEYFAPANTDANNIEGQAILYQ 2191
Qy      1383  -----GNWYYLDSNGEALTGSQRINDQRVFF 1408
Db      2192  NEFLTNGKKYYFGSDSKAVTGWRIINNKYYF 2224

RESULT 2
US-11-051-453-42
; Sequence 42, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMEROSINO, DONNA
; APPLICANT: BAECKOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-PEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: WJI-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 42
; LENGTH: 2367
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-42

```

2

Db 902 INKETSIFVETS---KTFSEYANHITBEISKIGTIFDVNGKL---VKKVNLDTTH 955
Qy 234 HNOIVNPE-----NTSLTIDNVVTADSWRPDKILKNGKTW-----TA 272
Db 956 EVNTLNAAFQOSLIENSSESLSNLSVAKVQVYAQLFSTGLNTTIDAAKVVELVSTA 1015
Qy 273 SSSES-DLRPLMSWPKQTOIAYLYNMYNOOGLGTGENYTADSSQESLNLAQAQVQVKIE 331
Db 1016 LDETIDLLPTLSEGLPIATII-----DGVSLGA-AIKELSETSDPLLRLQIEIAKI- 1065
Qy 332 TKISQTOOTOWLRDILINSFVKTPQNW-----NSQTESDTSAGEKH 372
Db 1066 -GIMAVNLTTATTAITTSGLASGFSILLVPLAGISAGISPLSVNNEVLVRDKATKVVDY 1124
Qy 373 LOGGALLYSN-----SDKTAYANSDYRLNRRPTTSOTGPKPYFEDNS-----S 415
Db 1125 FKHSVLETGVTLLDDKVMQDDLVI-----SEID-----FNNSIVLGKEIWRME 1174
Qy 416 GGYDFLLANDIDN--SNP-----VVQAEQLN-----439
Db 1175 GSGGHTVTDIDHFFSAPSITYREPHLSIYDVLEVQKEELDLSKOLMVLNPNRVPawe 1234
Qy 440 --W---LHLYMNGSIVANDPEANFDG-----VRVDAVD 468
Db 1235 TGWTPGLRSLENDGTLLDRIDNYEGEFYRWYFAFIADALITTLKPRYEDTINIRNLDS 1294
Qy 469 NVNADILQ-IASDYLK-----AHGVDKSEKN-AIN-HLSLEAWSNDPQYNK 514
Db 1295 NTKSFVITITTEIIRKUSYSPYGGGTVALPQLSQYNGMGINIELSEDVMI-----1346
Qy 515 DTKGAQLPIDNKLRLSLVALTRPLEKDAKNKEIRSGLEPVTITNSLNNRSAGKNSERM 574
Db 1347 -----IDVDNVVR-----DVTIESDKIKKGLIEGI-----LSTLSIE-----1379
Qy 575 ANYIFIRANDSEVQTVIAKIIKAQINPKTD--GLTFTL-----DELKQAFKI--- 619
Db 1380 ENKIIILNSHEIN-----FSGEVNGSGFVSLTFSILEGINAIIIEVDLLSKSYKLLIS 1431
Qy 620 -----YNEDMRQAKKY-----TQSNITPTAYA-----LMLSNDKS-----649
Db 1432 GELKIILMNSHIIQKIDYIGFNSQLQNIPIYFVDSGKENGFIINGSTKEGLFVSELDP 1491
Qy 650 ----ITRLYYGDMYSDGOYMATSPYYDAIDTLLKAKIKYAAG---GQDMKITY---VEG 700
Db 1492 VVLSKVYMDSDKSPFGY-----SNLKDVKVITKDNVAILTYLKDIIKISLSLTLOD 1547
Qy 701 DKS-----HMDWDTYGLTVSRVYGTGANEATDQGEATKTQGMVITSNPNSLK-----749
Db 1548 EKTIKLNSVHL--DESGVAEILKF-----MNRKSGSTNTSDLSMSFLESNN--IKSIFVN 1597
Qy 750 -LNQNDKVIWNMG-----AAHQKQYERPLLLITKDGTLSTYSDAAKSLY- 793
Db 1598 FLQSNIKFILDANFIISGTTSGQFEFICDENNNIQFYFIKFWLTETNYT-----LVV 1650
Qy 794 -RKTNDKGLVFPASDIQGLNPOVSGYLAVVPVCGASDNQDVRVAASKNANATGOVYES 852
Db 1651 GNRQNMIVPNVDLDP--SGDISSTVINFSQKYL-YGID-----SCVAKVISPNIY--1699
Qy 853 SSALDSQLIYEGFSNQDFVTKDSYTNKKIAQNVLQFKSWGVTSPFMAPQYVSSBDGS- 911
Db 1700 TDEINITPVYETNTVPEVILVDANYINEKINVIN-----DLSIRVWNSDND 1749
Qy 912 -----FLDSIIQN--GYAFEDRYDLAMSK-----933
Db 1750 FILMSTSEENKVSQVKIRFVNVFKOTLANKLSFNFSRQDQVPVSEIILSFTSPSYVEDGL 1809
Qy 934 -----NNKYGSOQDMINAVKALHKS-----GIQVIADWVD 964
Db 1810 IGYDLGLVSLYNEKFVYNNFGM---MVSGLIYINDSLYYFKPPVNNLITGFTVVGD---D 1863
Qy 965 QIYNLP---GKEVWTATRVNDYGEYRKDSEIKNTLYAANTKSGNDQYQAKYGAGFLSELA 1021

Db 1864 KYTFNPINGGAASIGETIIDDKNYFNQSGVLQT-----GVFSTEDG 1905
Qy 1022 AKYPSIFNRRTQIS-NGKKIDPSEKITAWKAKYFNGTNILRGVGYVLKDN---ASDKYFEL 1078
Db 1906 FKYPAPANTUDLENLEGAID-----FTGKLIIDENI--YYPEDNYRGAVWEKEL 1952
Qy 1079 KGNQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFYFDNNGHMVY 1138
Db 1953 DGEHMYFSPPE-TGK-AFKGLNQIGDDKYFNSDGMQK-GFV-SINDNKHVYFDDSGVMKV 2008
Qy 1139 GLQOLNGEVOYFTLSNGVOLRESFLENADGSKNTFGH-----LGNR-----YSNGYYSFDPN 1188
Db 2009 GYTEIDGKHFEYFAENG-EMOIGVFENTEDGFK-YFAHHNEDLGNBEGEIEISYS-GILFNFN 2065
Qy 1189 DSKWRVFDASGVMAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKKYFDGSGNMAVN- 1247
Db 2066 --KIYFDDSTAVGWKDL-----EDGSKYFFDEDTAEAYIGL 2102
Qy 1248 RFANDKNGWYILNSDGLALVGQTINGKTYFYFGQDGKQIKGK-----IITDNGKL- 1298
Db 2103 SLINDGQ---YYFNDGIMQVGFVTTINDKVYFSDSGIIESGVQNDIDNFIYIDNIGIVQ 2159
Qy 1299 -----KYFLANSDELARNIFAT-----DSQN 1319
Db 2160 IGVFTSDGYKYF-APANTVNDNIYQAVESGLVRVGEDVYFYFGETYTIETIETIYDWMEN 2218
Qy 1320 N-----WYFSGDGVAVTGSQTIAGKULYFASDQKQVKSFV 1356
Db 2219 ESDKYFVETKACKGINLIDDIKYFDEKIMRTGLISFENNYYFNENG-EIQFGYI 2277
Qy 1357 TYNGKVHYHADSGELQVNRFEADKDNWY-----LDSN--GEAL--TGSQINDORVF 1407
Db 2278 NIEDKMFYGED-GVMOIGVFNT-PDGFKYFAHQNTLDENFESESINTGLGLDEKRY 2335
Qy 1408 FTREGQVKGVDVAYD 1422
Db 2336 FTDEYIAATGSIID 2350

RESULT 3

US-11-052-554A-217
; Sequence 217, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 217
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-217

Query Match 5.0%; Score 370; DB 7; Length 726;
Best Local Similarity 27.0%; Pred. No. 2.8e-12;
Matches 110; Conservative 42; Mismatches 137; Indels 118; Gaps 10;
Qy 999 ANTKNGKDYQAKYGAGFLSELAAYKPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNI 1058
Db 67 AKTDSANKDPQKBTG-----SVATAPSMNSANNMSQSDKQNTVNEI-----108
Qy 1059 LGRGVGYVLKDNASKYFELKGNQTYLPK-OMTNKEASTGFVNDGNGMTFYSTGYQAKN 1117
Db 109 -----SSDSQQTKTDEQTLDPQNSFKQSAHVKMTTEAEKTFSHSI-----N 150

QY 213 TDTSOYQFKGLTKLNDYTPHNOIVNFENT--SLETIDNYVTADSWYRPDKILKNGKTW 270
Db 746 ADQSSRNIELGNTIQQDKKNAASINDILNTGFLKNNNNPIDFVSTYDIV-FANGNAT 804
QY 271 TASSEDLRPLLLMSWPDKOTQIAY-----LNMNQGL 304
Db 805 TATVTHDT-----ANKTSKVYDVNVDDTTIHLTGTDNKKLGKVTYKLNKTSANG- 855
QY 305 GTGENYVTADSSQESLNLAQOTVQVTKIETKISQTOQTQWLRI----- 346
Db 856 NTATNFVNSSDEDALVNAKDIAENLNLAKIHTTKGTADTALQTFVKKVDENNADD 915
QY 347 INSEVKTQPNWNSQTESDTSAGE-----KDLHGGALLYSNSDKTAYAN-----SDYRLLN 397
Db 916 ANAITVQCKNANKQNTLLTKGENGLNIKTDKNGVTFTGINTSGLKAGKSTLNDGGLSI 975
QY 398 RPTPTS-----QTGKP--KYFEDNSGGYDFLLANDIDNSNPVQAEQLNHLHLMYNGSIV 451
Db 976 KNPTGSEQIQVGADGVKFAKNNG-----VVGAGIDGTTITRDE-----IGFTGTNGSLD 1027
QY 452 ANDPEANFDGVRVDVNDVNNADLLQIASDYLKAHYG-----VDKSEKNA 495
Db 1028 KSRPHLSKDGINAGGKKTNIQSGEIAQNSHDAVTGGKIYDLKTELENKISSHTAKTAQNS 1087
QY 496 INHLSILEAWSN-----DPQYNKDTKGAQLPI-----DNKRLSLLYALTRPLEKDSAKN 547
Db 1088 LHEFSVADECGNFTVSNPYSSYDTSKTSDVITFAGENGITTKV-----NKG 1134
QY 548 EIRSGLE-----PVITNSLNKRSABGKNSERMANVIFIRAHDSVQTVIAKIIKAQIN 600
Db 1135 VVRVGIDQTKGLTTPKLTGNNN-----GKGI-----VIDSQNG 1168
QY 601 PKT-DGLTFTLDELKQAFKIYNEDMRQAKKKTQSNIPHTAYALMLNKSITRLYIGDMY 659
Db 1169 QNTITGLSNTL-----ANVTNDKGSVRTEQGNII 1198
QY 660 SDDGQYMAKSPYDAIDTLKARIKVAAGQDMKIYVEGDKSHMDWDYTGVLTSVRYG 719
Db 1199 KOEDKTRAA-----SIVDLSAGFNLQNGEAV-----DFVSTYDVNFA 1238
QY 720 TGANEATDQSEATKQGMVITSNPNLSKLNQNDKVIIVNMGAAHKNQEVRLPLLTQDG 779
Db 1239 DG-NATT-----AKTYDDTS-----KTSKVYDVNVDDTTIEVKDKKLGKVG-- 1279
QY 780 LYSYTSDAAKSLYRKTNB-KGELVFEDASDIQGLNPQVSGYLAVWVPVGCASDNQDVRVA 838
Db 1280 TTTLTSTGTGANKFALSNGATGALVKASDIVAHLN-TLGG-----DIQTA 1324
QY 839 ---ASNKANATQVYESSSALDSOLIIEGFSNFODFVTKDSDYTNKKIAQNVQLFKSWGVT 896
Db 1325 KGASQANNAGYV-----DADGNKVIYD-----STDNKY----- 1353
QY 897 SFEMAPQYVSSSEGSFLDSIIQNGYAFEDRYDLAMSKNNKYGQQDMINAVKALHKSIGQ 956
Db 1354 -----YQKNDGT-----VDKTEVAKDKLV 1374
QY 957 VIADWVPDQIYNLPGEVVVATRVNDYGBYRKDSKIKNTLYAANTKSNKGDYQAKYGGAF 1016
Db 1375 AQAQTPDGTGLAQMNVKSVINKEQVNDAN--KKQGINEDNAFVKGLEKASDNKTKNAVT 1432
QY 1017 LSEL--AAKPSIFNRQTSNGKKIDPSEKITAFAKAYFNGTNILGRGVGYVLKDNASDK 1074
Db 1433 VGDNLNAVQPTLTFAQDTGTAKKL--GETLTI-KGGQDNTNKLTDNNIGV-----AGTD 1485
QY 1075 YFELKGNQTYLPKQMTNKEA--STGPFVNDGNGMTFTYSTGYQAKNSFVQDAKGNWYFYN 1132
Db 1486 GETVK-----LAKDLTNLNSVAGGTTKIDDKGVSFVDSGQAKANTPVLISANG-----LDL 1536
QY 1133 NGHMV-----YGLQQLNGEVQYFLSNGVQLRESFLENADGSKNYFCHLGNRY 1180
Db 1537 GKQVLSNVKGTGKTDAAVNUQLN-EVRNLLGLG-----NAGNDNAGNQVNIADIKQDN 1591
QY 1181 NGYYS-----FDNDSKWRYPDASGVMAVGLKTINGNTQYFDODGY---QVK 1223

Db 1592 SGSSNRTVIAKAGTVLGGKGNNDTEKL---ATGGIQGV-----DKDGNANGDLS 1638
QY 1224 GAWI--TGSDDGKR-----YFDDGSGNMAVNRAN--DKNGDWYVYLNSDGIALVGVQTING 1275
Db 1639 NVWVKTQKDSKALLATYNAAGQNTYLTNNPABAIIR-----INBOGIRFFHVNDQNG 1692
QY 1276 KTYFYGQDGKQIKGKIITDNGKLYFLA-----NSGELARNIFATDSQNNWYFSGDGV 1329
Db 1693 EPVVQGRNG-----IDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNQSIAGDNA 1745
QY 1330 AVTGSQTI-----AGKKLYPASDQKQVGSFVTYNGKVHYTHADSGELQVNRFEADK 1381
Db 1746 QATGDQSIATGTGNVWAGKHSAGIDPSTVKAD-----NSYSVGNNGNOFTDATQTDV 1797
QY 1382 DG---NWYVLDNNGEALTSORINDQRVFFTFREGKQVK 1416
Db 1798 FGVGNNTVTTESNVALGSNSAISAG---THAGTQAK 1831
RESULT 6
US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: PROTEIN OF MORAXELLA
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-13
Query Match 3.8%; Score 280; DB 7; Length 1992;
Best Local Similarity 19.2%; Pred. No. 5.4e-07;
Matches 307; Conservative 206; Mismatches 603; Indels 482; Gaps 69;
QY 17 VTVAVASGLITLTGTTLGSVSASAEQTSQTSKDVVTVQKSEDDKA-----ASESSQT 66
Db 518 ISVDKSGNVTAPTYNIGVKTTELNSDGTSDKFSVKGSGTNNSLVTAHLASYLNEVNR 577
QY 67 DAPKTVQAOEQTOAQSOANVADTSTSIK-----ETPSONITTOANSDDKTVNTK 118
Db 578 -ADSALQSFVKEEDDDANAITVAKDTTKNAGVSIKLGKNGLTVATKKDQTVTFGL 636
QY 119 SBEAQTSSEERTKQSEBAQTASSQALTOAKAELTKQRTAAQENKQVLDLAIPIVVKQID 178
Db 637 SQDSGLTIKSTLNDGLTVKDTNEQIQVGANGIKFTNV---NGSNPG---TGIANARIT 691
QY 179 -GKYYYIGSDGQPKNFALTNNKVLVFDK-----NTG-----AL 212
Db 692 RUKIGFAGSDG-----AVDTNKPYLDDQKLVGNVKIITNGINAGKAITGSPILPSI 745
QY 213 TQTSQVQFKQGLTKLNDYTPHNOIVNFENT--SLETIDNYVTADSWYRPDKILKNGKTW 270
Db 746 ADQSSRNIELGNTIQQDKKNAASINDILNTGFLKNNNNPIDFVSTYDIV-FANGNAT 804
QY 271 TASSEDLRPLLLMSWPDKOTQIAY-----LNMNQGL 304
Db 805 TATVTHDT-----ANKTSKVYDVNVDDTTIHLTGTDNKKLGKVTYKLNKTSANG- 855
QY 305 GTGENYVTADSSQESLNLAQOTVQVTKIETKISQTOQTQWLRI----- 346

Db 856 NTATFNVNSDEDLVNAKDAENLNTLAKEIHTTKGTADTALQTFVTKKVDENNADD 915
QY 347 INSFVKTPQNNWSQTESDTSAGE-----KDLQGGALLYSNDKTAYAN-----SDYRLN 397
Db 916 ANAIVTGOKNANNQVNTLTKGEGNLNIKTDKNGTVTFGINTTSGLKAGKSTLNDGGLSI 975
QY 398 RPTPTS-----QTKGP--KYFEDNSSGGYDFLLANDINSNPVQAEQLNLHLMLMYGSIV 451
Db 976 KNPTGSEQIQVGADGVKFAKVNNG-----VWAGIDGTTTRDE-----IGFTGNGSLD 1027
QY 452 ANDPEANFQVRVDAVDNVDNADLLQIASDYLKAHYG-----VDKSEKNA 495
Db 1028 KSKPHLSKDGINAGGKKIINIOSGETAQASHDAVTGGKIKYDLKTELENKISSAKTAQNS 1087
QY 496 INHLSILEAWSN-----DPQYNKDTGAQLPT-----DNKLRLSLLYALTRPLEKDAASN 547
Db 1088 LHEFSVADEQGNFTVSNPYSSVDTSKTSDVITFAGENGITTKV-----NKG 1134
QY 548 EIRSGLE-----PVITNSLNNRSAGKNSERMANYIFIRAHSEVQTVIAKIIKAQIN 600
Db 1135 VVRVGIDQTKGLTTPKLTGVNNN-----GKGI-----VIDSQNG 1168
QY 601 PXT-DGLTFTLDELKQAFKIYNEDMEQAKKYTQSNIPYALMLSNKDSITRLYYGDMY 659
Db 1169 QNTITGLSNTL-----AKVTYDDTS-----KTSKVVDVNVVDVTIEVKOKKLGK-- 1279
QY 660 SDQGVYMATKSPYDAIDTLKARIKYAAGGDMKITYVEGDKSHMDWDYTGVLTSVRYG 719
Db 1199 KBEDKTRAA-----SIVDLVLSAGFNLQNGEAV-----DFVSTYDTVFA 1238
QY 720 TGAENATDOGEATKQGMVAVITSNNPSLKLNDQKIVNMGAHKNQYRPLLLTKDQ 779
Db 1239 DG-NATT-----AKVTYDDTS-----KTSKVVDVNVVDVTIEVKOKKLGK-- 1279
QY 780 LYSYTSDAAKSLYKRTND-KGELVPDASDIQGYLNPOVSGYLAVWVPVGSADNQDVRVA 838
Db 1280 TTTLTSTGTGANKFALSQATGDALVKASDIVAHLN--TLSG-----DIQTA 1324
QY 839 --ASKNANATGOVYESSSALDSOLIIEGFSNFQDFVTKDSYTNKKIAQNVLFKSWGVT 896
Db 1325 KGASQANNAGYV-----DADGNKVIYD-----STDNKY----- 1353
QY 897 SPFMAPQYVSSDESGFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVXALHSGIQ 956
Db 1354 -----YQAKNDGT-----DADGNKVIYD-----VDKTEVAKDKLV 1374
QY 957 VIADWVPDQIYNLPKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAP 1016
Db 1375 AQOATPDGTLAQNNVKSVINKEQVNDAN--KKQGINEDNAFVKGLEKASDNKTRNAAVT 1432
QY 1017 LSEL--AAKYPSTIFNRTOISNGKKIDPSEKITAWKAKYFNGTNILGRGVYVLKDNASDK 1074
Db 1433 VGDNLNAVATPLTFAGDTGTTAKKL--GETLTI-KGQYDNTNKLTDNNIGV-----AGTD 1485
QY 1075 YFELKGNQYLPKQWNTKEA--STGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFFN 1132
Db 1486 GFTVK-----LAKDLTNLSNVNAGGTKIDDKGVSVFVDSGQAKANTPVLISANG-----LDL 1536
QY 1133 NGHWV-----YGLQOLNGEVOYFPLNSGVQLRESFLENADGSKNYFCHLGNRYS 1180
Db 1537 GGVVISNVGKGTOKTDAANVQQLN-EVRNLLGLG-----NAGNDNADGNQNTIADIKDPN 1591
QY 1181 NGYYS-----FDNDSKWRYFDASGVMAVGLKTINGNTQYFPDQGY-----QVK 1223
Db 1592 SGSSSRNRTVIKAGTVLGGKGNNDTEKL--ATGGIQGV-----DKDGNANGDLS 1638
QY 1224 GAWI-TGSDGKKR-----YFDDGSGNMAVNRFAN--DKNGDWYLYNSDGLALVGQVOTING 1275
Db 1639 NVWVKTKDGSKALLATYNAAGQTNLYTNPAEADR-----INEQGIIRFFPHVNDGNG 1692
QY 1276 KTYVFGQDGKQIKGKIITDNGKLKYFLA-----NSGELARNIFATDSQNNWYFSGDGV 1329

Db 1693 EPPVQNRNG-----IDSSASGKHSVAIFQAKADGEAAVAIGRQTOAGNQSIAIGDNA 1745
QY 1330 AVTGSQTI-----ACKKLYFASDGKQVKGVSFTYNGKVVHYHADSGELQVNRPEADK 1381
Db 1746 QATGQDSIAIGTGNVAVAGKHSIGAIGDPSTVKAD-----NSYSVGNNNQFTDATQTDV 1797
QY 1382 DQ--NWYYLDSNGEALTCQSQRINDORVFFFTREGKQVK 1416
Db 1798 FGVGNNTVTESNSVALGNSAISAG-----THAGTOAK 1831

RESULT 7

US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PRP
; ORGANISM: Moraxella catarrhalis
US-11-013-759-4

Query Match 3.8%; Score 280; DB 7; Length 2047;
Best Local Similarity 19.2%; Pred. No. 5.6e-07;
Matches 307; Conservative 206; Mismatch 603; Indels 482; Gaps 69;

QY 17 VTVAVASGLITITGTTLGSVSAEATEQQTSDKVVTQKSEDDKA-----ASSESSQT 66
Db 573 ISVDAKSGNVTAPTYNIGVKTTELNSDGTSDKFSVKSGTNNSLVTAEHLASVLEVNRT 632
QY 67 DAPKTKAQTEQTAQSQANVADTSTSIK-----ETPQNITTOANSDDKTVTNYK 118
Db 633 -ADSALQSFVKEBEDDDANAITVAKDTTKNAGAVSILKLGKNGLTVA TKDGTVTFL 691
QY 119 SEEAQTSBERTKQSEBAQTASSQALTOAKAELTKORQTAQENKNPVDLAAIPNVKQID 178
Db 692 SQDSGLITGKSTLNDGLTVKQDTNEQIQVGANGIKFTNV---NGSNPG--TGIANAT 746
QY 179 -GKYYIGSDGQPKNFALTVNNKVLDFK-----NTG-----AL 212
Db 747 RDKIGFAGSDG-----AVDTNKPVLQDKLQGVNPKVITNTGINAGGKAITGLSPTLPSI 800
QY 213 TDTQVQFQKGLTKLNDYTPHNQIVNFPENT--SLETIDNYVTADSWYRPKDLLKNGKW 270
Db 801 ADQSSRNIELGNTIQDKDKSNAASINDILNTGFNLKNNNPIDFVSTYDIVD--FANGNAT 859
QY 271 TASSBSDLRPLLSMWPDQTOIAY-----LNYMNOOGL 304
Db 860 TATVTHDT-----ANKTSKVVDVNVVDVTIHLTGTDDNKKLGKVTKLKNTSANG-- 910
QY 305 GTGENYTDSSOESLNLAAQTQVQKIETKISQTOQTQWLRI----- 346
Db 911 NTATNFVNSDEDLVNAKDIAENLNTLAKEIHTTKGTADTALQTFVTKKVDENNADD 970
QY 347 INSFVKTPQNNWSQTESDTSAGE-----KDLQGGALLYSNDKTAYAN-----SDYRLN 397
Db 971 ANAIVTGOKNANNQVNTLTKGEGNLNIKTDKNGTVTFGINTTSGLKAGKSTLNDGGLSI 1030
QY 398 RPTPTS-----QTKGP--KYFEDNSSGGYDFLLANDINSNPVQAEQLNLHLMLMYGSIV 451

Db 1031 KNPTGSSQIQVAGDVKFAKVNNG-----VVGAGIDGTTTRTRDE-----IGFTGTNGSLD 1082
Qy 452 ANDPEANFDGVRVDAVONNADLLQIASDYLAHYG-----VDKSEKNA 495
Db 1083 KSKPHLSKOGINAGGKKTITNQSGLAQNSHDVATGGKIYDLKTELENKISSSTAKTAQNS 1142
Qy 496 INHLSILEASDN-----DPQNKDTKGAQLPI-----DNKURLSLLYALTRPLEKASKN 547
Db 1143 LHEFSVADEQGNFTVSNPYSSYDTSKTSVDITPAGENGITTKV-----NKG 1189
Qy 548 EIRSGLE-----PVLTSNNRSASGKNSERNYFIFRAHSEVQTVIAKLIKQIN 600
Db 1190 VVRGIDQTKGLTTPKLVGNNG-----GKGI-----VIDSQNG 1223
Qy 601 PKT-DGHTFTLDELKQAFKIYNEQMRQAKKKTQSNIPATAYALMSNKDSTIRLYYGDY 659
Db 1224 QNTITGLSNTL-----ANVNDKGSVRITTEQGNII 1253
Qy 660 SDDQYNATKSPYYDAIDTLKARIKYAAGQDMKITYVEGDKSHMDWDYTVGLTSVRYG 719
Db 1254 KDEDKTRAA-----SIVDLASGFNLQNGEAV-----DFVSTYDTVNPA 1293
Qy 720 TGANEATDQSEATKTQGMVITSNPSLKNQNDKIVNMGAAHQEYRPLLLTTKDG 779
Db 1294 DG-NATT-----AKVTYDDTS-----KTSKVYVDVNVDDTTIEVRDKKLGVK-- 1334
Qy 780 LTSVTSDAASLYRKTD--KGLVFDASDIOGLNPQVSGYLAVVVPVGSNDQDVRVA 838
Db 1335 TTTLTSTGTGANKFALSQATGDALVKASDIVAHLN-TLSG-----DIQTA 1379
Qy 839 --ASNKANATQOVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNVQLFKSGVGT 896
Db 1380 KGASQANNSAGYV-----DADGNKIYD-----STDNKY----- 1408
Qy 897 SFEMAPQYVSESGSFLDSLIQNGYAFEDRYDLAWSKNKYGSQQDMINAVKALHSGIQ 956
Db 1409 -----YQAKNDGT-----VDKTEKVAKDCLV 1429
Qy 957 VIADWYPDQIYNLPKGVEVTVATRVNDYGEYRKDSEIKNTLYAANTKSNQDYQAKYGGAF 1016
Db 1430 AQOATPDGTLAQMNVKSVINKEQVNDAN--KKQINEDNAPVKGLEKASDNTKNAVNT 1487
Qy 1017 LSEL--AAKYPSTIENRTQISNGKKIDPSEKITAWKAKYFNGTNIILRGVGYVLKDNASDK 1074
Db 1488 VGLDNVAOQPLTAFAGDTGTAKKL--GETLTI-KGGQDITNKLTDNNIGV--AGTD 1540
Qy 1075 YPELKGQDTYLPKQMTNKEA--STGFVNDGNGMTFYSTSYQAKNSFVQDAKGNWYFDW 1132
Db 1541 GFTVK-----LAKDLTNLNSVNAAGTKIDDKGYSFVDSGQAKANTPVLSANG-----LDL 1591
Qy 1133 NGHWV-----YGLQQLNGEVOYFSLNGVQLRESFLENADGSKNYFGLHGNRY 1180
Db 1592 GKGVISNVGKGTQDAAVQQLN-EVRNLLGLG-----NAGNDNAGNQVNIADIKKDPN 1646
Qy 1181 NGYYS-----FDNDSKWRYPDASGMVAVGLKTINGNTQYFDQDGY-----QVK 1223
Db 1647 SGSSNRTVIKACTVLLGKGNNDTEKL--ATGGIQGV-----DKDGNANGDLS 1693
Qy 1224 GAWI--TGSQDKR-----YFDDSGGMVAVNRPAN--DKNGDWTYLSGDIALVGVQTING 1275
Db 1694 NVWVKTKDQSKKALLATYNAAGQTNLTNNPAEADR-----INEQGIIRPFHVDNGQ 1747
Qy 1276 KTYFYQDQKQIKGKIITONGKLYFLA-----NSGELARNIFATDSQNNWYFSGDGV 1329
Db 1748 EPVVQGRNG-----IDSASGKHSVAIGFOKADGEAAVAIGROTQAGNQSIAIGDNA 1800
Qy 1330 AVTGSQTI-----AGKKLYPASDGKQVKGFSVTVNGKVHYHYHADSGELQVNRPEADK 1381
Db 1801 QATGDQSIAIGTVNVVAGHSGAIGDPSIVKAD-----NSYSVGNNNQFTDATQTDV 1852
Qy 1382 DG---NWWYLDNSGEALTGQSORINDQRPFTREGKQVK 1416
Db 1853 FGVGNITVTESNVALGSNSAISAG-----THAGTQAK 1886

RESULT 8

US-11-013-759-7
; Sequence 7, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-92LMIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-7

Query Match 3.8%; Score 280; DB 7; Length 2047;
Best Local Similarity 19.2%; Pred. No. 5.6e-07;
Matches 307; Conservative 206; Mismatches 603; Indels 482; Gaps 69;

Qy 17 VTVAASGLITLTGTLGSSVSAETEQTSDKVVTQKSEDDKA-----ASESSQT 66
Db 573 ISVDASGNTVATPTNYGVKTTLSNDSGTSDEKSVKSGTNNISLVTAELHSLVNEVNT 632
Qy 67 DAPKTYQAOEQEQAOASQANVADTSTSIK-----ETPSONITTOANSDDKTVNTK 118
Db 633 -ADSALQSEFTVKEEDDDANAITVAKDTTKNAGAVSILKLGKNGLTVAATKDDGTVTFGL 691
Qy 119 SEEAQTSERTKQSEEAQTASSQALTOAKAELTKQROTAAQENKQPVDLAAIPNVKQID 178
Db 692 SQSGUGITGKSTLNNDLGTVKDTNEQIQVGANGIKFTNV---NGSNPG--TGIANATARI 746
Qy 179 -GRYYYIGSDGQPKNFALTNNKVLVYFK-----NTG-----AL 212
Db 747 RDKIGFAGSDG-----AVDTNKPYLDQDKLVQGVNKKITNTGINAGGKAITGLSPTLPSI 800
Qy 213 TDTSTQYQKQGLTKLNDYTPHNQIVNFENT--SLETIDNYVTADSWYRPDKILKNGKTW 270
Db 801 ADQSSRNIELGNTIQDKKSNAAASINDILNTGFLNKNNNNPIDFVSTYDIDV-FANGNAT 859
Qy 271 TASSESDLRPLMSWHPDKQTQIAY-----LNYMNOOGL 304
Db 860 TATVTHDT-----ANKTSKVYVDVNVDDTTIHLTGDDNKKLGVKTKLKNKTSANG- 910
Qy 305 GTGENTADSSQESLNLAATQVQKLETQISQTOQTQWLRLDI----- 346
Db 911 NTATNFVNSSDEDALVNAKDAENLNTLAKEHTTKGTADTALQTFVTKVVDENNADD 970
Qy 347 INSFVKTPNNSQTESDTSAGE-----KDHLOGGALLYSNSDKTAYAN---SDYRLLN 397
Db 971 ANAIVTQXKANNQNVNTLTKGNGENLNIKTQKNGTVTFGINTTSGLKAGKSTLNDGGLSI 1030
Qy 398 RPTPS-----QTKGP--KYFEDNSSGGYDFELLANDINSNPVVQAEQLNHLHLYMVGSI 451
Db 1031 KNPTGSEQIQVAGDVKFAKVNNG-----VVGAGIDGTTTRTRDE-----IGFTGTNGSLD 1082
Qy 452 ANDPEANFDGVRVDAVONNADLLQIASDYLAHYG-----VDKSEKNA 495
Db 1083 KSKPHLSKOGINAGGKKTITNQSGLAQNSHDVATGGKIYDLKTELENKISSSTAKTAQNS 1142
Qy 496 INHLSILEASDN-----DPQNKDTKGAQLPI-----DNKURLSLLYALTRPLEKASKN 547
Db 1143 LHEFSVADEQGNFTVSNPYSSYDTSKTSVDITPAGENGITTKV-----NKG 1189

Db 1071 LESKINSAKTAQNSLHPSVADEQGNHFTVSNPYSYDTSKT--SDVITFA--GENGIT 1126
Qy 696 TYVEGDKSHMDWYTGVLTSVRVYGTGANBATDQGSBATKTQGMNAVITSNPNSKLQNOK 755
Db 1127 TKVNGVWRVIGIDQTKGLTPKLTUVGNNG--KGIVDSKQDQNTITGLSNTLANVTND- 1183
Qy 756 VIVNMGAAHKNQYRPLLLTKDGLTSYSDAAKSLYKRTNDKGLVFDASDIQGYLNP 815
Db 1184 -----GAGH-----ALSOGLANDTDKTRAASI-----GDTLANAGFNLOQ--NG 1219
Qy 816 QVSGYLAVVPVVGASDNQDVRVAASNKANA--TCQVYESSALDSQLIYEGFSNFQDFVTK 874
Db 1220 EAVDFVSTYDVFID-----GNATTAKVYDTSKTSKVY----- 1256
Qy 875 DSDYTNKKIAQNVOLFPMKSGVTSFEMAPQVYSSSEDSGLDSIIQNGYAFEDRYDLAWSKN 934
Db 1257 DVNVNDKTI--EVTSDKGLGVKTTLTK--TSANGNATKFSAADGALVKASDIATHLN 1311
Qy 935 NKYGSOQDMINAVKA-----LHKSIGIQTADVVPQIYNLPKGEVVTATRVNDYGEYRK 988
Db 1312 TLAGDIQTAKGASQASASSAYVDADGNKVIYDSTDKKY-----QVNDKGQVDK 1360
Qy 989 DSEI-KNTLYAANTKNGKDYQAKYGAFLSELAAYPSIFNRTOISNGKKIDPSEKITA 1047
Db 1361 NKEVAKOLVA-----QAQTPDGTIAQNVK--SVINKEQVND-----A 1397
Qy 1048 WKAKYFNGTNILGRGVGYVLKONASDKYFELKGNQTYLPKQMTNKEASTGFVNDGNGMTF 1107
Db 1398 NKKQGINEDNAFTKIGLENAADTKTKNAAVTVGDLNAVAQTPLTFAGDTGTTAKKLGETL 1457
Qy 1108 YSTSGYQAKNSFVQDAGKNWYFFDNNGHMYGLQQLNGEVOYPLSN-----GVOLRE- 1159
Db 1458 TIKGGQTDNKL-T-----DNNIGVVAGTGTFTVLAKDLTNLNSVNAAGGTIDEK 1507
Qy 1160 --SFLNADG-----SKNYFGLHGNRYSN-GYVSPDND--KWRYPDASGVMAVGLK 1206
Db 1508 GISFVD-ANGQAKANTPVLNGLDILGKKRISNIGAAVDNDNAVNFQKNEVA-----K 1560
Qy 1207 TINGNTQYFDQGVQV-----KGAWITGSDGKKRYFDDGSGNMVNRFPANDKNGDWY 1259
Db 1561 TVN-NLNNQNSGASLPFVVTDANGKPIGTGDKPQKAIGA-----DGKYYH 1607
Qy 1260 LNSDGLALGVQTINGKTYFGQDGKQIKKIITDNGKLYFLANSSEL--ARNIATDS 1317
Db 1608 ANANGVPV-----DKDGRP-----ITDADKLANLAAGKPLDAGHQVVASLG 1649
Qy 1318 QNNWYFGSDGVAVT 1332
Db 1650 GN-----SDAITLT 1658

RESULT 11
US-11-052-554A-231
; Sequence 231, Application US/11052554A
; Publication No. US2005028886A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 231
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-231

Query Match 3.6%; Score 270; DB 7; Length 701;
Best Local Similarity 19.9%; Pred. No. 4.4e-07;
Matches 178; Conservative 114; Mismatch 285; Indels 318; Gaps 37;
Qy 470 VNADLLQIASDYLKAHYGVDKSEKNAIHLSTILEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529
Db 89 LNIKLSAITKYLRL-----ELNVLE-----EKSD-----ELPSEIKAKL 123
Qy 530 SLLYA-----LTPPLEKASNKNEIRSGLEPVITNSLNRSAEGKNSEMANIYIFIRAH 584
Db 124 DAAFEKFKDULPKGEKVAEAKKVEEA-----KKKAEDOKEDRRNY-----PT 168
Qy 585 SEVQTVIAKIIKAOINPKTDGLTFLDELKQAFKIYNE--DMROAKKYTQSNIPATAYALM 643
Db 169 NTYKTELEIEAFDVKVKEAELELVKEEAKES---RNEGTTIKQAKEK-----V 213
Qy 644 LSNKDSITRLYYGDMYSDDGQVMATKSPYDAIDTLLKARIKYAAGQDMKITVYVBGDKS 703
Db 214 ESKKABATRL-----ENIKTRKKAEEAKKAKADAKL----- 245
Qy 704 HMDWYTGVLTSVRVYGTGANEAAT--DQGSBATKTQGMNAVITSNPNSKLQNOKDQVIMMGA 762
Db 246 -----KEANVATSDQ-----KPKGRA----- 262
Qy 763 AHKNOBYRPLLLTTKDGLTSTYSDAAKSLYKRTNDKGLVFDASDIQGYLNPQVSGYLA 822
Db 263 ----KRGVFGELATPD-----KKEND----- 279
Qy 823 VWPVPGASDNQDVRVAASNKANATQVYESSALDSQLIYEGFSNFQDFVTKDSYTNKK 882
Db 280 -----AKSDSSVGEETLPSLSLKGKVAEAEKKEEAEKAKAD----- 319
Qy 883 IAQNVOLFPMKSGVTSFEMAPQVYSSSEDSGLDSIIQNGYAFEDRYDLAWSKNKYQSOD 942
Db 320 ---QKEEDRRNYPTNTYKTLDLLEIAESDVKVKEAELE-----LVKEEAKPRDEE 366
Qy 943 MINAVKALHKSQIQTADVVPQIYNLPKGEVVTATRVNDYGEYKDSIKNTLYAANTK 1002
Db 367 KIKQAKAKVES-----KKAEATRLNLIKTRKKAEE-----AKRK 402
Qy 1003 SNGKDYQAKYGAFLSELAAYPSIFNRTOISN--GKKTDPSEKITAWKAKYFNGTNILG 1060
Db 403 AAEBEK-----VKEKPAEQPAPATQPKPAPKPEKPAEQPKAEKT----- 444
Qy 1061 RGVGYVLKDNASDKYFELKGNQTY--LPKQK-----TNKEA-----STGFVNDGNGMTFYST 1110
Db 445 -----DQQAEDYARRSEBEYNRLTQQQPKTEKPAQSPKTPKTGMQENGMMWYFYNT 497
Qy 1111 SGYQAKNSFVQDAGKNWYFFDNNGHMYGLQQLNGEVOYFSLNGVQLRSFLENADGSKN 1170
Db 498 DGSMA--TGWLQ--NGSWYILNANGAMATGWLQNGSWYIL-----NANGS-- 540
Qy 1171 YFGHLGNRYSGVYSPDNDKWRYPDASGVMAVGLKTINGNTQYFPDQD-----YQVK 1224
Db 541 -----MATGW--LQNGSWYILNANGAMATGWLQNGSWYILNANGAMATGWLQYNG 590
Qy 1225 AWITGSDGKKRYFDDGSGNMVNRFPANDKNGDWYILNSDGLIALVGQTINGKTYFGQDG 1284
Db 591 SW-----YYLANGDMATGWLQN--NGSWYILNANGDMATGWLQNGSWYILNANG 639
Qy 1285 KQIKGIITDNGKLYFLANSSELARNIFATDSQNNWYFFGSDGVAVTCSQITAG 1339
Db 640 DMATGWV--KDGDTWYILEASGAMKASQWFKVS--DKWYVNGSG--ALAVNTTVDG 690

RESULT 12
US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shiesong, Jiang

;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
;; FILE REFERENCE: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
;; CURRENT APPLICATION NUMBER: US/11/022,562
;; PRIOR FILING DATE: 2004-12-22
;; PRIOR APPLICATION NUMBER: PCT/US03/20322
;; PRIOR FILING DATE: 2003-06-27
;; PRIOR FILING DATE: 2003-06-27
;; PRIOR FILING DATE: 2002-06-27
;; NUMBER OF SEQ ID NOS: 340
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 228
;; LENGTH: 396
;; TYPE: PRT
;; ORGANISM: Clostridium difficile
US-11-022-562-228

Query Match 3.6%; Score 267.5; DB 7; Length 396;
Best Local Similarity 26.7%; Pred. No. 2.8e-07;
Matches 94; Conservative 45; Mismatches 124; Indels 89; Gaps 16;
QY 1051 KYFNGNIIIRG-VGVVLKDNASDKYFELKGNQYLPKQMTNKEASTGFVNDGNGMTFFYS 1109
DB 91 KSFENSELDRDLHLGFKIDNKTYYYDEA-----SKLVKGLININNSLFYFD 137
QY 1110 -----TSGYQAKNSFVQDAKGNWYFD--NNGHMVYGLQQLNGEVOYFLSNGVQLRESEL 1162
DB 138 PIESNLVTGWQTN-----GKKYFIDINTGAASTSYKIINGKHFFYFNNGV-MQLGVF 189
QY 1163 ENADGSKNTFGLHGNRYN-----GYYSFDNDSKRWYFDASGVNAVGL 1205
DB 190 KGPDPGE-YFAPANTQNNIEGQAIYVQSKFLTLNGKTYFNDNSK-----AVTGW 239
QY 1206 KTINGNTQYFDQGVQVKGAWITGSGKKRYFPD-----GSGNMAVNRFANDKNGDWYLL 1260
DB 240 QTIIDGKYYFNLTAEAATGWQT-IDGKKYFNTNTSIASGTII-----NGKHFFY 291
QY 1261 NSDGIALVGV-QTINGKTY--FGQDGKQIKGKIITDNGKLYFLANSGLARNIPATDS 1317
DB 292 NTDGIWQIGVFGPGNGFEYFAPANTDANNIEGQAIYRQNFYLY----- 335
QY 1318 QNNWYFSGDGVAVTGSQIAGKLYFASD-GKQVKGSTFYNGKVHYTHAD 1368
DB 336 HDNIYFNGNSKAVTGWQTINGVYFMPDTAMAAAGGLFEIDGVYIFFGVD 387

RESULT 13
US-10-793-626-2964
;; Sequence 2964, Application US/10793626
;; Publication No. US20050255478A1
;; GENERAL INFORMATION:
;; APPLICANT: KIMMERLY, WILLIAM JOHN
;; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;; FILE REFERENCE: PU3480US
;; CURRENT APPLICATION NUMBER: US/10/793,626
;; CURRENT FILING DATE: 2004-03-04
;; PRIOR APPLICATION NUMBER: 60/164,258
;; PRIOR FILING DATE: 1999-11-09
;; NUMBER OF SEQ ID NOS: 4472
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 2964
;; LENGTH: 5024
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: synthetic
;; OTHER INFORMATION: amino acid sequence
;; FEATURE:
;; NAME/KEY: MOD RES
;; LOCATION: (5024)
;; OTHER INFORMATION: variable amino acid
US-10-793-626-2964

Query Match 3.6%; Score 266; DB 6; Length 5024;
Best Local Similarity 18.9%; Pred. No. 1e-05;
Matches 259; Conservative 211; Mismatches 561; Indels 340; Gaps 59;
QY 18 TVAVASGLITLGTITLGSSVSAETEQTSDKVVTO-----KSEDDKAAESSOTDA 68
DB 1060 TINQLANAVTQAKSNLHGDTKLQHDKDSAKQTIAQLQNLNSAQKHMEDSLIDNESTRT-- 1117
QY 69 PKTKQAQTEQTOAQS-----QANVADTSTSTIKETPS-----QN 102
DB 1118 ----QVQHDLTEAQALDGLMGALKESIKONTNIVSNGNYINAE-PSKKQYDAVQVQAQN 1172
QY 103 ITTQAN-----SDDKTVTNTKS-----EEAQTSEERTQS-----BEAQTAS 140
DB 1173 IINGTNOPTINKGNVTTATQTKNTKDALDGDHRLSEAKNNANQTIIRNLNNAQKDAE 1232
QY 141 SQALTOAKA-ELTKQRTAAQENKVPD--LAAIPNVKIDGKYIYIGSGQPKNF--- 194
DB 1233 KNLVNSASTLEQVQQLQTAQQLDNAMGBLRQSIANKQOVKADSKYLNEDPOIKQNYDDA 1292
QY 195 -----ALTVNKVLVYFDKNTGALTDTDSQYQFKOGLTKLNNDYTPHNQIVNPFNTSLE 246
DB 1293 VQVRETIINETQNPPELLKANIDQATOSVQNAEQALHGAELNOD-----KQTSST 1342
QY 247 TIDNYVTADSWYRPKDILKNGKWTWASSESRLRPLMSWMPDKQTQIAYLN----- 297
DB 1343 ELDGLTDLTDAQREK--LREQINTNSRDDIK-----QKIEQAKALNDAMKKLKEQ 1391
QY 298 YNNQGLGTGENYNTADSS--QBSLNLAQTVQVKIETKISQTOOTQWLRIIINSFVKTP 355
DB 1392 VAQKQGVHANSYDTNEDSAQDAYNNALKAQEDIIINNSNPNLNAQDITNALNNIKQAQD 1451
QY 356 NNNS-----QTESDT--SAGEKDH--OGGALLYSNSDKTA----- 387
DB 1452 NLHGAQKQODKNTNTQAIQNLNLNQPOKDALIQAINGATSQDVAEKLKEAEALDEAM 1511
QY 388 ----YANDYRLLANRP--TSQTGPKYFENSSGGYDFLLANDN--SNPVOAEQL 438
DB 1512 KQLEDQVNDQDQISNSSPINEDSDKQTYNDKIQ-----AKEIINTSNTLQKQKI 1565
QY 439 -----NWLH-----YLMYVGSIVANDPEANFGV--RVDADVNVNA 472
DB 1566 ADTLQNIKDAVNNLHGDQKLAQSKODANNQNLHLDLTBEQKNHFKPLINNADTRDEVNK 1625
QY 473 DL--LQIASDYLKAHYGVDKSEKNAIHL--LLEAWSNDNDPOYNKDTKGAQLPIDNKL 528
DB 1626 QLEIAKQLNGDMSTLHKVI--NDKQOIQHLISNYINADNDKKQNYDWAKEAEOLIHHPD 1683
QY 529 LSLLYALTRPLEKADASNKEI--RSGLEPVTINSLNRSAGKNSRMANYIFIR---AH 583
DB 1684 TLDHKALQDLNLKIDQAHNELNGESRFKQALDNALND--IDSLSLNVPQRTVKDNINH 1741
QY 584 DSEVOTVIAKITIAQINPKTDLGTLTFLDELKQAFK-----IYNEDMRQAKKYQTSNI-- 636
DB 1742 VTTLESQAQELQAK-----ELNDAMKWRDSIMNQEIRKNSNYTNEDLAQ 1788
QY 637 PTAYALMLSNKDSITRLYYGDMYSDGQYMAKSPYDAIDTLLKARIKYAGGQDMKIT 696
DB 1789 QNAYNHAVDNIINNII-----GEDNATMDPQI-----IKQATQDINTATN 1827
QY 697 YVEGDKSHMDMDYTGVLTSVRYGTGANEATDQ-----SEATKTOGMVITSNPSEL--K 749
DB 1828 GLNGDQKLQD-AKTDAKQOITNFTGLTPEKQALENIINQOTSRANVAKQLSHAKFLNGK 1886
QY 750 LNQNQDKVINMGAAHKNQYERPLLTTKDGLTSYTSDAAKSLYRKTNDKGLVFDASDI 809
DB 1887 MEELKVAVAKASLVRQNGNY-----INEDVSEKAEYEAIAKQGEIINSEN- 1932
QY 810 QGYLNPQVSGYLAVWVPVQASDNQDVRVAASKANATQGVYESSALDSQLI-----Y 862
DB 1933 ----NPTIS-----STDINRTIQEINDAEQNLHGENKLROAQEIAKQNIQNL 1975
QY 863 EGFNSFQDFVTK---DSDYTNKKIAQNVQLPKSWGVT-SFEMAPQVVSSEDSGFLDSIIQ 918

Db 1976 DGLNSAQ--ITKLIQIDGRTTKPAVTQKLEBAKAINQAOQLKQSIADKDATLNS-- 2030
QY 919 NGYAFEDRYDLAMSKNNKYGSDQDMINAVK--ALHKSIGQIVADWPDQIYNLPGKEVVT 976
Db 2031 -NYLNEDS-EKKLAYDNVAVSQAEQLNLQNLNDPTMDISNIOAITQKVIQAKDSLHG---- 2083
QY 977 ATRVNDYGEYRKDSEIKNTLYAANTKSNKQVQA-----KYGGAFLSELAAKYPSIFN 1029
Db 2084 ---ANKLAQOADS---NLIIQSTNLNDKQKQALNDLHQAQTKQOVAEIIAQANKLNN 2137
QY 1030 RTQISNGKKIDSEKIKATWAKYFN-----GTNILGRGVGYVLKON-ASD 1073
Db 2138 --EMGTLKTLVBEQSNVHOQSKYINEDPQVQIYNDISIQKGREILNTTDDVLNNKIKAD 2195
QY 1074 KYFELKGNQTYLTPKQMTNKEASTGFVNDGNGMTFFYSTGYQAKNSFVQDAGNWWYFDNN 1133
Db 2196 AIQNIHLTKNDLHGDKLQKQAQDATNELNLTNLNLSQKQSEHDEINSAPSRTEVSNDL 2255
QY 1134 GH---MVYGLQQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSN 1181
Db 2256 NHAKALNEAMROLENEVA--LENSVRKLSDFINDEAAQ-----NEYSN 2297

RESULT 14

US-11-052-554A-229
; Sequence 229, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 229
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-229

Query Match 3.6%; Score 265; DB 7; Length 619;
Best Local Similarity 19.8%; Pred. No. 6.8e-07; Indels 244; Gaps 31;
Matches 158; Conservative 106; Mismatches 300; Indels 244; Gaps 31;
QY 642 LMLSNKDSITRLYGDYSDGYMATKSPYDAIDTLKARIKYAAGQDMKITYVEGD 701
Db 6 MILTSLASVAILGAGFVASQPTVRAEESP-----VASQSKAEKDYDAAKKDAK---NAK 57
QY 702 KSHMDWDTYGLTVSRYGTGANFATQGSSEATTKQGMVITSNPGLKNDNDKVINMG 761
Db 58 KAVED-----AQALDDAKAAQK-----KYDEDQKTEEXA 88
QY 762 AAHK--NOEYRPLLLTTKDLGTSY--TSDAASAKSLYRKNTDKG-----ELVFPASDIQY 812
Db 89 ALEKASEEMDKAAVAQYQAYLQOATDIAKADKADKIDEAKKREBEAKTFNVRAM 148
QY 813 LNPQVSGYLAVWVPVGASNDQDVRVAASNKA-NATQGVYESSALDSQLIYEGFSNFQDF 871
Db 149 VVPE-----PEQLAETKKSEAKQKAPELTKLBEAKALEE----- 186
QY 872 VTKDSYTNKIAQNQLFKSWGVTSFEMAPQVVSSE-----DGSFLDSI 916
Db 187 -----AEKKATEAQK-----KVDAAEVAPOAKIAELENOVHRLBEQLKEIDSESESDY 234
QY 917 IONGY--AFEDRYDLAMSKNNKYGSDQDMINAVKALHKSIGQIVADWPDQIYNLPGKEV 974

Db 235 AKEGFRAPIQSLDKAKAKLSKLELSDKIDELDA-----EIAK-LEDQL-----KAA 281
QY 975 VTATRVNDYGE-----YRKDSEIKNTLYAANTKSNKQVQYQAKYGGAFSELAAKYPSI 1027
Db 282 EENNVEDYFKEGLEKTTIAAKKAELEKTEADLKCAVNEPEKPAPEPETPAPEAPAEQPKP 341
QY 1028 FNRTQISNGKKID-PSEKITAWKAKYFNGTNIILGRGVGYVLKDNASDKYFELKGNQTY-- 1084
Db 342 APAPQAPAPKPEKPAEPKPEKT-----DDQQAEDYARRSEBEYNR 384
QY 1085 LPKQMTNK-----EASTGFVNDGNGMTFFYSTGYQAKNSFVQDAGNWWYFDNNHMYV 1138
Db 385 LQOQPPPPKAEKPAPEKPTGWKQNGWYFYNTDGSMA-TGWLQN-NGSWYILNSNGAMAT 442
QY 1139 GLQQLNGEVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSGYSPNDKSKWRYFDAS 1198
Db 443 GWLQYNG-----SWYILNAN 457
QY 1199 GYMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWY 1258
Db 458 GANATGWAKVNGSWYILNANGAMATG-WL-----QYNGSWY 492
QY 1259 YLNSDGIALLVGQTINGKTYFYFGDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQ 1318
Db 493 YLNSANGAMATGWAKVNGSWYILNANGAMA-----TGMLOY-----N 528
QY 1319 NNWYFPGSDGVAVTGTSQTIAGKLYFASDGKQVKSFTYTYNGKVHYHADSGLQVNRPE 1378
Db 529 GSWYILNANGAMATGWAKVNGSWYILNANGAMATG--WYKDGDTWYILEASGAMKASOWP 586
QY 1379 ADKDGWYILDSNGEALTGSRINDQV 1406
Db 587 KVSD-KWYIVNGLG-ALAVNTTVDGKYV 612

RESULT 15

US-10-873-528-17
; Sequence 17, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hanabro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT FILING DATE: 2004-06-23
; PRIOR FILING DATE: 2004-06-23
; PRIOR FILING DATE: 2001-01-26
; PRIOR FILING DATE: 2001-01-26
; PRIOR FILING DATE: 1998-03-27
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-17

Query Match 3.6%; Score 265; DB 6; Length 658;
Best Local Similarity 24.1%; Pred. No. 7.4e-07; Indels 106; Gaps 21;
Matches 106; Conservative 59; Mismatches 168; Indels 106; Gaps 21;
QY 1048 WKAKYF---NGTNILGRGVGYVLKONASDKYFELKGNQTYLTPKQMTNKEASTGFVNDGNG 1104
Db 28 WQCKQYLKEDGGQAANEV---FDTHYSWFYIKADANY-----AENELKQDGD 74
QY 1105 MTFYSTYSQYQAKNSFVQDAGNWWYFDNNHGM-----VTGLQ- 1141
Db 75 YFLKSGGYMAKSEWED-KGAFYILDQDGKMKRNAWVGTSYVGATCAKVIEDWYDSQY 133

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 29.5158 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-36

Perfect score: 7462

Sequence: 1 METKRYKMKVKHWTVA.....EGKQKGVAYDERRLLVYR 1430

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	7443	99.7	1431	2 A45866	dextranucrase (EC
2	3817.5	51.2	1475	2 B3135	gtfB protein precu
3	3788.5	50.8	1577	2 T30858	glucosyltransferas
4	3713.5	49.8	1449	2 T30857	glucosyltransferas
5	3713.5	49.8	1449	2 T30552	glucosyltransferas
6	3591.5	48.1	1375	2 JT0345	dextranucrase (EC
7	3542	47.5	1592	2 A38175	glucosyltransferas
8	3502.5	46.9	1518	2 A44811	glucosyltransferas
9	3312	44.4	1599	2 S22737	glucosyltransferas
10	3252.5	43.6	1508	2 T31098	probable dextranu
11	3244.5	43.5	1290	2 JCS473	dextranucrase (EC
12	3174	42.5	1365	2 A41483	glucosyltransferas
13	562	7.5	2817	2 B97033	uncharacterized pr
14	451	6.0	2710	2 A37052	toxin A - Clostrid
15	435	5.8	563	2 A37184	glucan-binding pro
16	412	5.5	2178	2 S55805	alpha-toxin - Clos
17	393	5.3	2367	2 S70172	toxin B - Clostrid
18	388	5.2	2364	2 A10884	cytotoxin L - Clos
19	387	5.2	2366	2 S10317	toxin B - Clostrid
20	363.5	4.9	648	2 S10869	enterotoxin A - Cl
21	306.5	4.1	1463	2 T30290	AAS surface protei
22	296.5	4.0	690	2 F98114	choline-binding pr
23	293.5	3.9	2334	2 S32920	cell wall-associat
24	291	3.9	1335	2 T30211	autolysin B - Stap
25	278	3.7	721	2 C97980	endo-beta-N-acetyl
26	275.5	3.7	6713	2 B89921	hypothetical prote
27	272	3.6	1819	2 D97033	uncharacterized pr
28	272	3.6	1881	2 H95076	zinc metalloprotei
29	270	3.6	701	2 H98120.	choline binding pr

30	268.5	3.6	1946	2 AEL1449	hypothetical prote
31	265	3.6	619	2 A97887	surface protein ps
32	265	3.6	619	2 A41971	surface protein ps
33	265	3.6	627	2 G97975	hypothetical prote
34	265	3.6	658	2 E95111	hypothetical prote
35	263	3.5	2167	2 AFI1489	endo-beta-N-acetyl
36	262	3.5	601	2 S57962	cell wall-associat
37	262	3.5	3890	2 C89921	cspC protein - Clo
38	261	3.5	1385	2 D89824	hypothetical prote
39	260.5	3.5	2285	2 T12796	hypothetical prote
40	259	3.5	1296	1 HMS01P	probable transglyc
41	257.5	3.5	340	2 G95043	aggregation protei
42	257.5	3.5	627	2 E95107	choline binding pr
43	257.5	3.5	2271	2 F90073	hypothetical prote
44	255.5	3.4	1423	2 A49206	exo-beta-D-fructos
45	255.5	3.4	1959	2 AG1085	hypothetical prote

ALIGNMENTS

RESULT 1

A45866

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004

C:Accession: A45866

R/Honda, O.; Kato, C.; Kuramitsu, H.K.

J. Gen. Microbiol. 136, 2099-2105, 1990

A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl

A:Reference number: A45866; MUID:91100958; PMID:2148600

A:Accession: A45866

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1431 <HON>

A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296

C:Keywords: glycosyltransferase; hexosyltransferase

F:181-201/Domain: cpl repeat homology <CP1>

F:1127-1146/Domain: cpl repeat homology <CP2>

F:1192-1211/Domain: cpl repeat homology <CP3>

F:1257-1276/Domain: cpl repeat homology <CP4>

F:1277-1297/Domain: cpl repeat homology <CP5>

F:1321-1340/Domain: cpl repeat homology <CP8>

F:1341-1361/Domain: cpl repeat homology <CP6>

F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 99.7%; Score 7443; DB 2; Length 1431;

Best Local Similarity 99.9%; Pred. No. 7e-318;

Matches 1427; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	METKRYKMKVKHWTVA	VAVASGLITLGTTLGSSVSAETBOOTS	SDKVVTQKSEDDKAA	60
DB	1	METKRYKMKVKHWTVA	VAVASGLITLGTTLGSSVSAETBOOTS	SDKVVTQKSEDDKAA	60
QY	61	SSSSQTDAPKTKQAQTEQ	TQAQSQANVADTSITKETPSQNIIT	QANSDDKVTNTKSE	120
DB	61	SSSSQTDAPKTKQAQTEQ	TQAQSQANVADTSITKETPSQNIIT	QANSDDKVTNTKSE	120
QY	121	EAQTSEERTKQSEEAQT	TASSQALTOAKAELTKQRQTAAQENKN	PVDLAAIPNVKOIDGK	180
DB	121	EAQTSEERTKQSEEAQT	TASSQALTOAKAELTKQRQTAAQENKN	PVDLAAIPNVKOIDGK	180
QY	181	YYIIGSDGPKKNFALT	VNNKVLVFDKNTGALTDSYQYQKGL	TKLNNDYTPHNOIVNF	240
DB	181	YYIIGSDGPKKNFALT	VNNKVLVFDKNTGALTDSYQYQKGL	TKLNNDYTPHNOIVNF	240
QY	241	ENTSLETIDNYTADSW	TRPKDILKNGKTTWTSSESDLRPLLS	WWPDKOTQIAYLNYN	300
DB	241	ENTSLETIDNYTADSW	TRPKDILKNGKTTWTSSESDLRPLLS	WWPDKOTQIAYLNYN	300
QY	301	QQGLGTGENYTADSS	QESLNLAATQVQVKIETKISQ	QQTQWLRLDINSFVKTPNNSQ	360
DB	301	QQGLGTGENYTADSS	QESLNLAATQVQVKIETKISQ	QQTQWLRLDINSFVKTPNNSQ	360

```
QY 361 TESDTAGKDHLOGGALLYSNDSKTYANDSYRLNRTPTSTQTKPKYFEONSQGYDF 420
Db 361 TESDTAGKDHLOGGALLYSNDSKTYANDSYRLNRTPTSTQTKPKYFEONSQGYDF 420
QY 421 LLANDIDNSNPVQAEQLNLWHLMYGSIIVANDPEANFDGVRVADVANNADLLQIASD 480
Db 421 LLANDIDNSNPVQAEQLNLWHLMYGSIIVANDPEANFDGVRVADVANNADLLQIASD 480
QY 481 YLKAHYGVDSKSEKNAIHNLSILEAWSNDPOYNKOTKGALPLIDNKLRLSLLYALTRPLE 540
Db 481 YLKAHYGVDSKSEKNAIHNLSILEAWSNDPOYNKOTKGALPLIDNKLRLSLLYALTRPLE 540
QY 541 KDNASNKNEIRSGLEPVITNSLNRSABGKNSERMANIYIFIRAHDSVQTVIAKIIKAQIN 600
Db 541 KDNASNKNEIRSGLEPVITNSLNRSABGKNSERMANIYIFIRAHDSVQTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPATYALMLSNKDSITRLYYGDMYS 660
Db 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPATYALMLSNKDSITRLYYGDMYS 660
QY 661 DGQYMATKSPYYDAIDTLLKARIKAYAGQDMKIITYVEGDKSHMDWDYTGVLTSVRYGT 720
Db 661 DGQYMATKSPYYDAIDTLLKARIKAYAGQDMKIITYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQSEATKTQGMVITSNPESLKLNDQKVIYVNGAAHKNQOEYRPLILTTKQGL 780
Db 721 GANEATDQSEATKTQGMVITSNPESLKLNDQKVIYVNGAAHKNQOEYRPLILTTKQGL 780
QY 781 TSVTSDAAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLAWVPVGSADNDQVRVAAS 840
Db 781 TSVTSDAAKSLYRKNTDKGELVFDASDIQGYLNPQVSGYLAWVPVGSADNDQVRVAAS 840
QY 841 NKANATGQVYESSALDSQLIYEGFNFQDFVTKDSYTNKKIAQNVLFKSNGVTSFEM 900
Db 841 NKANATGQVYESSALDSQLIYEGFNFQDFVTKDSYTNKKIAQNVLFKSNGVTSFEM 900
QY 901 APOYVSEDSFLDSIIQNGYAFEDRYDLAMSNNKYGSQDDMINAVKALHSGIOVIAD 960
Db 901 APOYVSEDSFLDSIIQNGYAFEDRYDLAMSNNKYGSQDDMINAVKALHSGIOVIAD 960
QY 961 WVPDQIYNLPGEKWTATRVNDGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
Db 961 WVPDQIYNLPGEKWTATRVNDGEYRKDSEIKNTLYAANTKSNKGDKYQAKYGGAFSEL 1020
QY 1021 AAKYPSIFNRTOISNGKKIDPSEKITAWKAKYPNGTNILGRGVYVLDKONASDKYFELKG 1080
Db 1021 AAKYPSIFNRTOISNGKKIDPSEKITAWKAKYPNGTNILGRGVYVLDKONASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVODAKGNWYFDNNGHMYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGNGMTFYSTGYQAKNSFVODAKGNWYFDNNGHMYGL 1140
QY 1141 QQLNGEVQVFLSNGVOLRESFLENADGSKNYFCHLGNYSNGYYSFNDNSKRYFDASGV 1200
Db 1141 QQLNGEVQVFLSNGVOLRESFLENADGSKNYFCHLGNYSNGYYSFNDNSKRYFDASGV 1200
QY 1201 MAVGLKTINGTOYFPDQGVQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
Db 1201 MAVGLKTINGTOYFPDQGVQVKGAWITGSDGKKRYFDDGSGNMAVNRFANDKNGDWYYL 1260
QY 1261 NSDGIALVGQTINGKTYFQDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
Db 1261 NSDGIALVGQTINGKTYFQDQKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WYFVSGDGVAVTGSQITAGKLYFASDGQVKGSVFTYNGKHYYHADSGELQVNRFEAD 1380
Db 1321 WYFVSGDGVAVTGSQITAGKLYFASDGQVKGSVFTYNGKHYYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYLYDSNGBALTGSQRINDQVFFTRREGQVKGDVAYDERRLLVY 1429
Db 1381 KQGNWYLYDSNGBALTGSQRINDQVFFTRREGQVKGDVAYDERRLLVY 1429
```

RESULT 2

B33135 gtfB protein precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004

C:Accession: B33135; A33128

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013; PMID:3040685

A:Accession: B33135

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1475 <SH1>

A:Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN

R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

submitted to the Protein Sequence Database, September 1990

A:Reference number: A33128

A:Accession: A33128

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-171,173-641,'N',643-1475 <SH2>

A:Cross-references: UNIPARC:UPI000017AC5E

A:Experimental source: strain GS-5

F:1096-1115/Domain: cpl repeat homology <CP1>

F:1224-1243/Domain: cpl repeat homology <CP2>

F:1289-1308/Domain: cpl repeat homology <CP3>

F:1354-1373/Domain: cpl repeat homology <CP4>

F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match

Best Local Similarity 51.2%; Score 3817.5; DB 2; Length 1475;

Matches 762; Conservative 230; Mismatches 395; Indels 53; Gaps 22;

```
QY 1 METKRYKMKHKVKGHWTVVAVASGLITLCTTTTGGSSVSAETBQSTQSDKVVTKSEDDKAA 60
Db 1 MDKVKYKLRKVKRWTVVSVASAVMTL--TTLSGGLVKADSNESK-----QISNDNSIS 54
QY 61 SSSQTDAPKTQAOETEQAQOASQAN--VADTSTSTTKETPPSONITTOANSDDKTVTNTKS 119
Db 55 VVTANESNVI TEATSKQEAASSQTNHTVTTSSTSTSVVNPKEVSNPYTVGETASNGEK 114
QY 120 EBAQTEERTKQSEBAQTASSQALTQAKAELTKQRTAAQENKQNPVDLAALPNVKQIDG 179
Db 115 LQNTQTTV--DKTSEAAANNISKQT--TEADTVDDSNAA----NLQILEKLPNVKAEIDG 167
QY 180 KYIYIGSDGQPKNFALTVNNKVLVFDKNTGALTDTTS--QYQPKQGLTKLNNDYTPHNQIV 238
Db 168 KYIYDNGKVKRTNLTIAADGKILHFE--TGAYTDTSIDTVNKDITVTRSNLYKKYQVY 226
QY 239 NFENTSLETIDNYVTADSWYRPKDILKNGKWTWASESDLRPLLLMSWPDKQTOIAYLNY 298
Db 227 DRSAQSFHVDHYLTAESWYRPKYILKDGKWTQSTEKDFRPLMTWPDQETQOYVNY 286
QY 299 MNQOGLGTCENTADSSQESLAAQTVQKLETKISQTOQTCWLRLDIINSFVKTPQNN 358
Db 287 MNAQ--LGINKTYDDTSNQLNIAAATIOAKIEAKITTLKNTDQLRTQISAFVKTSQAMN 345
QY 359 SOTESDTSAGEKHLOGGALLYSNSDK--TAYANSDYRLNRTPTSTQTK--PKYFEDNS 415
Db 346 SDEKPPD----DHLQNGAVLYDNEGKLTPIYANSVYRLNRTPTQTKKDPRTADNTI 401
QY 416 GGYOFLANDIDNSNPVQAEQLNLWHLMYGSIIVANDPEANFDGVRVADVANNADLL 475
Db 402 GGYEFLANDVDNSNPVQAEQLNLWHLMYGSIIVANDPEANFDGVRVADVANNADLL 461
QY 476 QIASDYLKAHYGVDSKSEKNAIHNLSILEAWSNDPOYNKOTKGALPLIDNKLRLSLLYAL 535
Db 462 QIAGDYLKAAKGIGHKNDKAANDHLSILEAWSNDPTPLHDDGNNMINDKRLSLFLSL 521
QY 536 TRPLEKDAENKNEIRSGLEPVITNSLNRSABGKNSERMANIYIFIRAHDSVQTVIAKII 595
```


Db	939	EGFSNFQDFVKTPSQYNRVIAQNAKLFKEWGITSEFPAPQVSSQDGFSLSIENGVA	998
Qy	923	FEDRYDLAMSKNNKYQSQDMINAVKALHKSQIQTADWVDPQIYNLPKGEVVVATRVND	982
Db	999	FEDRYDIAMSKNNKYSGLKDLMDALRALHAEGISAIADWVDPQIYNLPKGEVVVATRTNS	1058
Qy	983	YGEYRKDSSEIKNTLYAANTKYSNGKDYOAKYGGGAFSELAAKYPISIINRQISNGKKIDPS	1042
Db	1059	YGTPEPNAIYNSLYAAKTRTFGDNDFQGGYGGAFDELKAKYPAIFERYQISNGRKLTTN	1118
Qy	1043	EKITAWKAKYFNGTNTLGRGVGVYVKDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDG	1102
Db	1119	EKITQSAKYFNGSNLOGTGARVLODNATNQYFSVKAQTFLPKQMTITGS-GRPRVG	1177
Qy	1103	NGMTFYSTSGYQAKNSFVQDAKGNWYYPDNNGHMYGLQOLNGEVOYFSLNSGVQLRESFL	1162
Db	1178	DDVOYLSIGYLAKNTFIQVGANOWYYPDKNGNMVTGEQVIDGKYFFLDNGLQLRHVLR	1237
Qy	1163	ENADGSKNYFPHLGNRYSGYISFPDNDK-WRYFDASGMAVGLKTINGNTQYFDQ-DGY	1220
Db	1238	QSGDGHVYYPDKGVQAFNGFYDFAGPRQDVFYFDGNGQMYRGLHDMYGTTFYFDESKTI	1297
Qy	1221	QVKGAMITSGDKRKYFDDGSGNMAVRFA-NDKNGDWYLLNSDGLALGVQOTINGKTY	1279
Db	1298	QAKDKFIRFADGTRFIPDTGMLAVNRFAQNPENKAWYLDNSGVAVTGLQTINGKQY	1357
Qy	1280	FGQDGKQIKITDNGKLKYL-ANSSELARNIFATDSQ-----	1318
Db	1358	FDNEGRQVKGHFTVINNO-RYFLDGSDEIAPSRFVTENKNWYVDGNGKLVKAQVING	1416
Qy	1319	-----NNWYFSGDGVAVTGSQTIAG	1339
Db	1417	NHYFNNYDSQVKGAWANGRYDGDGSAQVSNQFIQIAANOWAYLQDGHKVTGLQINN	1476
Qy	1340	KKLYFASDGVQKVGSPVYNGKVHYHADSGELQVNRFEADKGNWYLDNSGEALTGSQ	1399
Db	1477	KVYFSGNAQVKGKULTVQGGKCYFDAHTGEOQVNRFEAARGCWYFNSAGQAVTGOQ	1536
Qy	1400	RINDQRVFFTRREGQVKGVDAVY	1421
Db	1537	VINGKQLYFDGSGRQVKGKRVY	1558
RESULT 4			
T30857			
glucosyltransferase I- Streptococcus salivarius			
C:Species: Streptococcus salivarius			
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T30857			
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.			
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri			
Infect. Immun. 63, 609-621, 1995			
A:Reference number: Z20909; MUID:95122197; PMID:7822030			
A:Accession: T30857			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1449 <SIM>			
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI0000081666; EMBL:L35495; NID:g662378; PII			
C:Genetics:			
A:Gene: gtfL			
Query Match 49.8%; Score 3713.5; DB 2; Length 1449;			
Best Local Similarity 51.4%; Pred. No. 9.8e-155;			
Matches 754; Conservative 207; Mismatches 409; Indels 97; Gaps 22;			
Qy	1	METKRYKQHKVKKHVTAV-----ASCLITLGT-----TTLGSSVSAET	41
Db	1	MDKKVYKQHKVKKQWVTAVTGLSLGAVSAVSLGTNGDGVQVADEHTATVAIPDITVDI	60
Qy	42	EQOTSQKVVTKSEDDKAASESQTD--APKTKQAQTEQTQAQSQANVADTSTSI-----	94
Db	61	GTVSNUTTAQAQDPTTAVATNDVATQATPTATFDLTDTTNTVAANAVDTAVATVGTDR	120

Qy	95	--TKETPSQNIIT--QANSDDKTVNTKSEEAQTSERTKQSEEAQTASSQALTOAKAE	150
Db	121	ATTNDTATNDTAVDTNNNTTDTTVDRAATERRATGARRGPT--GGRRTPVNGN	178
Qy	151	LTKQRTAAQENKNPVDLAAIPNV-----KQIDGKYIYIGSDGQPKNFALTVNNKV	202
Db	179	TNNANTVTVVNN---DLPATNNVTDGFSHLKTINGKQYVVEDDGTIRKNVLERIGS	235
Qy	203	LYFDKNTGALTPTSQYQF--KQGLTKLNNDYTPHNOIVN-----FENTS---LETIDNYVT	253
Db	236	QYFNAETGELSKEQYRFDKNGCTGSSADSTNTNTVNGDKNAFYGTDTDKDIELVDGYET	295
Qy	254	ADSWRPDKILKNGKWTATASSEDLRPLLMSPWPKQTOIAVLYNMNOGLGTGENYTAD	313
Db	296	ANTWYRPKEILKDGKEWTASTENDKRPLLTVMWPSKAIQASLYNMKEGLGTNGTYTTF	355
Qy	314	SSQESLNLAQTVQVKIETKISQTOQWMLROIINSFVKTPQNMNSQTESDTSAGEKHL	373
Db	356	SSQTDQDALEVKRIERIAREGNTDWLRTTIKNFVKTPQWNSTSE---NLDNNDHL	412
Qy	374	QGGALLYNSDKTAVANSDYRLNLRTPPTSQTK--PKYFEDNSSGGYDFLLANDIDNSNP	431
Db	413	QGGALLYNDSRTSHANSDYRLNLRTPPTSQTKHNPKYTKDTSNGGFEFLANDIDNSNP	472
Qy	432	VYQAEQLNWLHLYMNYGSIIVANDPEANFDGVRDAVDVNVNADLLQIATSYLKAHYGVDS	491
Db	473	AVQAEQLNWLHYIMNIGTITGGSSEDEFDGVDAVDVNVNADLLQIATSYLKAHYGVDS	532
Qy	492	EKNALNHLISLEAWSNDPQYNKDTKGAOLPIDNKLRLSLLYALTRPLEKDNKNKIEIRS	551
Db	533	QQQALKHLISLEAWSNDAYNEDTKGAOLPMDPDMHLALVLSLLRPIGN-----RS	584
Qy	552	GLEPVTISLNNRSRSGKNSERMANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFTLD	611
Db	585	GVEPLISLNSLDRSESGKNSKRMANYAFVRAHDSVQSIIGLIKNEINPQSTGNTFTLD	644
Qy	612	ELKQAFKIYNEDMQAKKYITQSNIPTAYALMNSKDSITRLYXGDMYSDGQYMATKSP	671
Db	645	EMKKAFEIYNKDMRSANKQYTOYNIPSAVALMTHKQDTPRVYVGYDMYTDGQYMAQKSP	704
Qy	672	YVDAJDTLLKARIKYAAGGQDMKITVVEGDKSHMDWDTGVLTSVRYGTGANEATQDGS	731
Db	705	YVDALETLLKGRIRYAAGQDMKNVYI--GYGTNGHDAAGVLTSVRYGTGANSASDTGTA	763
Qy	732	ATKTQGMVITSNPNSLKNQNDKIVNMGAHKKQYERPLLLTTKDGITSYTSDDAAKS	791
Db	764	ETRNQGMVIVSNQPALRLTSN--LTINNGAAHRNQAYRPLLLTTNDGVATVYLNDSANG	821
Qy	792	LYRKTNDKGELVFDASDIQYLNPOVSGYLAVWVPVVGASDNQDVRVAASNKANATQVYE	851
Db	822	IVKYTDGNGNLTFSAINEIRGNPQVDGYLAVWVPVVGASENQDVRVAPSKENSSGLVYE	881
Qy	852	SSSALDSQLIYEGFSGNFQDFTKDSYTNKKAQNVOLFESKSWGTSEFEMAPQVSSSEDS	911
Db	882	SNAALDSQVIYEGFSGNFQDVFQNPQSYTNKKIAENANLFKSWGITSFEFAPQVSSDDGS	941
Qy	912	FLDSIQNGYAFEDRYDLAMSKNNKYQSQQDMINAVKALHKSQIQTADWVDPQIYNLP	971
Db	942	FLDSVIQNGYAFDTRYDIQMSKNNKYGLADLKAALKSLHAVGISAIADWVDPQIYNLP	1001
Qy	972	KEWVATRVNDYGEYRKDSEIKNTLYAANTKNGKDYQAKYGGAFSELAAKYPISFNRT	1031
Db	1002	DEVVATRVNNYGETKDGAIIDHLSYAAKTRTFGNDYQYQYGGAFDELKRLYPPQIFDRV	1061
Qy	1032	QISNGKKIDPSEKITAWKAKYFNGTNTLGRGVGVYVKDNASDKYFELKGNQTYLPQMTN	1091
Db	1062	QISTGKRWTTDEKITQWSAKYNGTNIIDRGSEYVLK--NGLNGYYGTNGKVSPLPKVGS	1120
Qy	1092	KEASTGFVNDGNG-----MTFYSTSGYQAKNSFVQDAKGNWYFDDNGHMYGLQ	1141
Db	1121	NQSTNGDNQNGDGSQKFKRLFSVRVRYNNGYQAKNAFIKNDNGVYVFDNSGRMAVGEK	1180

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1592 <ABO>

A;Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:G217032; PIDN:BAAL4241.1; PID:
F:1093-1112/Domain: cpl repeat homology <CP1>
F:1222-1241/Domain: cpl repeat homology <CP2>
F:1287-1306/Domain: cpl repeat homology <CP3>
F:1330-1351/Domain: cpl repeat homology <CP4>
F:1352-1371/Domain: cpl repeat homology <CP5>
F:1402-1420/Domain: cpl repeat homology <CP6>
F:1465-1484/Domain: cpl repeat homology <CP7>
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 47.5%; Score 3542; DB 2; Length 1592;

Best Local Similarity 48.0%; Pred. No. 3.6e-147;

Matches 714; Conservative 242; Mismatches 411; Indels 120; Gaps 27;

```
QY 1 METKRYKMKHKVKKHVVAVASGLITLGTITGSSVSASBTEQOTSDKVVTKSEDDKAA 60
DB 1 MEKNVRFKMKHKVKKRWVTLVSASA--TMLASALGASVAS-----ADTDAS 44

QY 61 SESSQTDAPKTKQAQTEQTAQSQANVADTSTSTTETPSONITTOANSDDKTVNTKSE 120
DB 45 DSNQOT---VVTGDTTNNQATQOTSTAAATATS-EGSASTDAATDQASAEQOTQGTAST 100

QY 121 EAQTSBERTKQSEBAQ---TTASSQALTOAKAELTKQRTAAQENKPNVDLAAIPNVKQI 177
DB 101 D--TAAQTITNANEAKWVPTENENQGTDEMLAEAKNVAT-AESDSISPDLAKMSNVKQV 157

QY 178 DGKYYIGSDGQPKFPALTVNNKLVYFDKNTGALTDTSTQYQFQKGLTKLND---YTPH 234
DB 158 DGKYYIYDQGNVKNKFAVSGVKIYYFDG-TGAYKDTSDVDADKSSVASQATIPAAAN 216

QY 235 NQIVNFENTSLETIDNVVTADSVRPKDIKNGKTKWTASSESLRPLLMKSWPDKTOIA 294
DB 217 NRAYSTSAKNFEAVDNYLTADSVYRPKASILKQKTKWESGDDFRPLMAWVPTETKRN 276

QY 295 YLYNMNQGLGTGENTYDASSQBSLNLAAQTVQVKIETKISQOTQOTQWLRIINSFVKIQ 354
DB 277 YVNMNKKV-VGIDKTYTAETASQADLTAAAEVLQARIEQKITSENNTKWLREALSAFVKIQ 335

QY 355 PNWNSPTSEDSAGEKHLOGGALLYSN-SDKTAYANSYRLNLRPTSQG-----KPKY 409
DB 336 PQMGSESEPYD----DHLQNGALLFDNQDITLFTQSNRYLLNLRPTQNTQSLDSRFTY 391

QY 410 FEDNSSGGYDFLLANDIDNSNPVVAEQNLNHLVLMNYSIVANDPEANFDGVRVDAVN 469
DB 392 NPNDPLGGYDFLLANDVDNSNPVVAEQNLNHLVLMNYSIVANDADANFDSIRVDAEN 451

QY 470 VNADLLQIASDYLKAHYGVDKSEKNAINHLSSILEAWSNDNDPQYNKDTKGAQLPIDNKLRL 529
DB 452 VDADQLQISSDYLKAAYGIDKNNKNANNHVSIVAEWSNDNTPYLHDDGDNLMNMNKNFRL 511

QY 530 SLLYALTRPLEKDAKSNKIRSGLEPVINSLNRRSAEKNRSMANYIFIRAHDSVQT 589
DB 512 SMLWSLAKP-----TDVRSGLNPLIHNLSVDREVDDREVEVTPSYSFARAHDSVQD 563

QY 590 VTAIKIKAQINPKTDCGLTTLDELKQAFKLYNEDMQAKKYTOSNIPTAYALMLSNKDS 649
DB 564 ITRDIKABINPNSFGYSTQEBIDQAFKIYNEDLKKSKDKYTHYNVPLSYTLTLNKGSS 623

QY 650 ITRLYYGDMSYDDGQYMATKSPYDAIDTLTKARIKYAAGGQDMKITTYVEGDKSHMDWDY 709
DB 624 IPRVYVYGDMTDDGQYMANKYVNYDAIESLLKARMYVAGQAMQ-NYQINGE----- 676

QY 710 TGVLTSVRYGTGANETDQGSSEATKTOGMNAVITSNPNPSLKLNDKQVIYVMGAHKNQY 769
DB 677 --ILTSVRYGKALKQDKGDAITRTSGVGVWGNQPNFSLD-GKVVALNMGAHAHQY 733

QY 770 RPLLLTKDGLTSYTSDDAAK--SLYRKTNDKGELVFDASDITQGLYNPQVSYLAWVWY 827
DB 734 RALMVSTKDGVAITYATDADASKAGLVKRTDENGILYFLNDDLKGVANPQVSGFLQVWV 793
```

```
QY 828 GASDNQDVRVAASNAKANATQVYESSSALDSQLIYEGFSNFQDFVTTKDSYTNKKIAQNV 887
DB 794 GAADDQDIRVAASDSTASTDCSKSLHODAAAMDSRVMEFEGFSNFQSPATKEEYTNVVIANNV 853

QY 888 QLFKSWGVTSFEMAPQVYSSDSFLDSIIQNGYAFEDRYDLAMSKNKYGSGQODMNAV 947
DB 854 DKFVSWGITDFEMAPQVYSSDQFQDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAI 913

QY 948 KALHKSQIOVADWPDQIYNLPKQEVVTVATRVNDVGEYRKDSSEIKNTLVAAANTKSGKD 1007
DB 914 KALHAKGLKVMADWVPDQMTFFPKQEVVTVTRTDKFGKPIAGSOINHSIVTDTKSSGDD 973

QY 1008 YQAKYGGAFSLAELAAKYPISIFNRTOISNGKKIDPSEKITAKKAYFNGTNIILGRGVYVL 1067
DB 974 YQAKYGGAFSLDELKQYKPELFTKKQWSTGQAIIDPSVKIKQWSAKYFNGSNILGRGADYVL 1033

QY 1068 KDNASKYFELKGNQYTLPKQMTNKAESTQFVNDGNGMTFYST-SGVQAKNSFVQDAKGN 1126
DB 1034 SDQVSNKYFNVASDITLFLPSSLGLGKVVEGIRYDGGKIYINSSATGQDVKASFITEA-GN 1092

QY 1127 WYFVDMNGHMVYGLQQLNGEVQVFLSNGVOLRESFLENADGSKNYFGLGNRY--SNGYY 1184
DB 1093 LYTFGKGDMVTGQATNGANYFFLENGTALRNTIYTDAGNSHYVANDGRKVENENGYQ 1152

QY 1185 SPDNDKRWYFDASGVMAVGLKTINGTQYFDODGQYQVGAWITGSDGKRYFDDGSGNM 1244
DB 1153 QFGND--WRYF-KDGNMVGTLTVGDNVQYFDKDGQVQAKDKIIVTRDGKRYFDDHNGNA 1209

QY 1245 AVNRPANDKNGDWYILNSDQIALUGVQTINGTQYTFGQDQKQIKGKIIT-DNGKLKYLFLA 1303
DB 1210 VTNFTIADTGHWHYLGKGVAVTGAQTVGKQKLYFEANGQYQVKGDFVTSDEGKLYFYDV 1269

QY 1304 NSGELARNIPATDSQNNWYVFGSDGVAVTGSGQTIAGKLLYFASDGQVKSFTV-YNGKV 1362
DB 1270 DSGDMWTFIEDKAGNWFILGKGAATVGAQTIRGOKLYFKANGQQVKGDIKVGTDGKI 1329

QY 1363 HYHADSGELQVNRFEADKDG----- 1383
DB 1330 RYDYDAKSGQVFNKTKVKAADGKTYVINGDVADVPVVKVQTFKDSAGALRFYNLKGQLV 1389

QY 1384 -----NWYILDNSGEALTGSRINDQVRVFTREGKQVKGDV 1419
DB 1390 TSGMWYETANHDWVYIQS-GKALTGQTINGOHLFYFKDGHQVKGQL 1435

RESULT 8
A44811
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C;Species: Streptococcus salivarius
C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: A44811; S27296; S28809
R;Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A;Title: Molecular characterization of a cluster of at least two glucosyltransferase ge
A;Reference number: A44811; MUID:92148377; PMID:1838391
A;Accession: A44811
A;Molecule type: DNA
A;Residues: 1-1518 <GIF>
A;Cross-references: UNIPROT:Q00600; UNIPARC:UPI000008BEF31; EMBL:Z11873; NID:947526; PID
A;Note: sequence extracted from NCBI backbone (NCBI:81050, NCBI:81052)
C;Genetics:
A;Gene: gtfJ
C;Keywords: glucosyltransferase; hexosyltransferase
P;1307-1326/Domain: cpl repeat homology <CP4>
```

Query Match 46.9%; Score 3502.5; DB 2; Length 1518;
Best Local Similarity 48.1%; Pred. No. 1.8e-145;
Matches 724; Conservative 247; Mismatches 434; Indels 101; Gaps 32;

```
QY 1 METKRYKMKHKVKKHVVAVAS-GLITL--GTTILGSSVSASBTEQOTSDKVVTKSEDD 57
DB 1 MENKIYKHLKVKQWVITAVASVALATVGLGLSVTTSSVSAD---ETQDKTVTQ-SNSG 56
```


Db 168 DN1K--KEKDKYKYLLEDGSHKQNFALTNGQVLYFDEN-GALSTSTSYSPQTETNLV 224
QY 229 NDYTPHQIYNFENTSLTIDNYVTADSWYRPKDILKNGKTTWTASSSDLRPLMSWPD 288
Db 225 TDFTKNAAVDSTKASPELVGVLTDADSWYRPKEILLEAGITWKASTEKDFRPLMSWPD 284
QY 289 KQTQIAYLNTNQQGLGTGEN----YTADSSQBSLNLAAQTQVQVKEITKISQTOQTOWLR 344
Db 285 KDTQVAYLNTMT-KALNSGEBETKDVFTIENSQASLAAAOILQRIKIEVKIAANKSTDWLR 343
QY 345 DIINSFVKTOPNNSOTESDTSAGEKHLQGGALLYSNDSKTAAYANSDYELLARTPTSQT 404
Db 344 QSTEAFVKDQKNWINSERP-----GKEHFKGALLFVNSDSTKWSNDYRKLNQATSYI 399
QY 405 GKPKYFEDSSGGYDFELLANDIDNSNPVQAEQLNHLHLYLNTYGSIV--ANDPEANFDGV 462
Db 400 KHKH-IVNGSDGGYEFLLSNDIDNSNPVQAEMLNQLYFPMWQIVFGDKDAHFDGI 458
QY 463 RVDADVNNADILQIASDYLKAHYGVDKSEKNAINHLSILEANSDDNPQYKDKTKGAQLP 522
Db 459 RVDADVNSVDMQLQVSSYMKAAKYVNESEARALANISILEANSNDPPYVNEBHNTAALS 518
QY 523 IDNKLRLSLLYALTRPL-----EKDASNKEIRSGLEPVIITNSLNNRSAEKNSERM-- 574
Db 519 MDNGLRLSVHGLTRPVTNKGTRGARNASMKDLINGGY-----FGLSNR-AEVTSDQLGF 572
QY 575 ANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFLDELKQAFKIYNEDMROAKKYTOS 634
Db 573 ATYLFVRANDSEVQTVIADIISKIDPTDGTFTLDQLKQAFDIYNADMVKDKYTHIS 632
QY 635 NIPTAVALMSKDSITRLYYGDMYSDGQYMATKSPYDAIDTLKARIKYAAGQDMK 694
Db 633 NIPAAVALMLQTMGAATRVYGLYTDNGQYMAKSPYFQITLLKARPKYVAGQTSY 692
QY 695 IITVEGDKSHMDWYTGVLTSVRYGTGANEATD-QGSEATKQGMVAITSNPSKLQN 753
Db 693 IHNLAGDVSSAKDNKEVLVRYGQDLMSKTDTEGGYGRNSGMLTLIANNPDLKADG 752
QY 754 DKVIVNMGAAHKNQRYRPLLLTKDGLTSYTSDAARSLYRKNTDKGELVFQASDIQYL 813
Db 753 EITVNMGAHKNQARPLLLGTEKIVSSLNDSDTK-IVKYTDAQGNLVFTADEIKGPK 811
QY 814 NPQVSGYLAVWPVPGASDNQVRVAASNKANATG-QVYESSALDSQLIYEGFSNQDFV 872
Db 812 TVDMSGYLSVWPVPGATDQNVLAKPSTKAYEGDKVYSSAALAEQVIVEGFSNFQDFV 871
QY 873 TKDSYTNKLIQNVQLFKSNGVTSSEMAPOYVSSSEDSGLSDIIONGVAFEDRYDLAMS 932
Db 872 KEDSQYTNKLIANADLFKSWGITSFEIAPQYVSSKDGTFLLDSIIENGVAFTRDYDFAMS 931
QY 933 KNNKYSQODMINAVKALHKSQIQLVADWVPQIYNLPGKEVVVATRVNDYGEYRKDSI 992
Db 932 KNNKYSKEDLRDALKALHKGQIQLVADWVPQIYLPKGEVVVATRTDTHGKVLDDTSL 991
QY 993 KNTLYAANTKSGKQYQAKYGAFSLSELAAKYPSIFNRTQI1SGNKKIDPSEKITAWKAY 1052
Db 992 VNKLXYNTKSGNDFOAQYGAFLDKLQKLYPEIFKEVMEASGKTIDPSVKIKWEAKY 1051
QY 1053 FNGTILGRGVYVLKDNADSKYFELKGNQTVLPKQMT-NKEASTGFVNDGNGMTPYSTS 1111
Db 1052 FNGTNIQRKGSYVLSB--GKLYFTVNDKGTFLPALTGDTRAKTGFAYDGTGVITYTTS 1109
QY 1112 GYQAKNSFVODAKGNWYVFDNNGHMVYGLQQLNGEVQYFSLNGVQLRESFLENADGSKNY 1171
Db 1110 GTQAKSQFV-TYNGKQYFNDKGILYVTGSEQTDIGSNYFFLPNGWMTDGRNKNAGQSULV 1168
QY 1172 FGHGNNR-----YSGNY--- 1183
Db 1169 YKSGKLTQTGKWEVTVKDSGKBEKFYQYFFKGGIMATGLTEVEGKEKYFDNGYQAK 1228
QY 1184 -----YS--FDNDSKWRVFDASGVMAVGL----- 1205
Db 1229 GVFVPTKDHLMFFCGDSGERKYSGFQEDGNWYANDKGYVATGFTKVGKQNLVFNK 1288

QY 1206 -----KTINGNTQYFDDGYOVYKAWITGSDGKKRYFD 1238
Db 1289 VOYKNRFFQVDATYANNEGDVLRAQTINGDELIFDESGKQVKGEFVANNPDOTTSSYD 1348
QY 1239 DSGGNMAVNR-----FANDKNGDWYLLNSDGIALVGVQTIIN 1274
Db 1349 AITGVKLVTSLVVVDGQTFNVDAKVGVVTKAHTPGFYTTGDNNNFYADSYGRNVTTGAQVIN 1408
QY 1275 GKTYEYFGQDKQIKGLIITD-NGKLXYFLANSSELARNIIFATDSQNNWYFSGDGVAVTG 1333
Db 1409 GQHLFYDANGRQVKGFVNTDGSRSFYHWNTGDKLVSTFFATGHDHRYVYADDRGNVVTG 1468
QY 1334 SQTIAKGLYFASDGQVKVGSFVT-YNGKVHYTHADSGELQVNRFEADKGNWYLLDSNG 1392
Db 1469 AQVINGOKLFFDIDGQVKGAFATNANGRSYTHWNTGNKLVSTFFTSQNNWYADAKG 1528
QY 1393 EALTGSQRINDQVRFFTRGKQVKGVA 1420
Db 1529 EVVVGQTINGQHLYPDQTKQVKGATA 1556

RESULT 10
T31098
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31098
R:Monchois, V.; Renaud-Simeon, M.; Monean, P.; Willemot, R.M.
FEMS Microbiol. Lett. 159, 307-315, 1998
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (D
A:Reference number: Z20981; MUID:98164374; PMID:9503626
A:Accession: T31098
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1508 <MON>
A:Cross-references: UNIPROT:O52224; UNIPARC:UPI00000BB69B; EMBL:AF030129; NID:g2766611;
A:Experimental source: strain NRRL B-1299
C:Genetics:
A:Gene: derB
C:Function:
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 43.6%; Score 3252.5; DB 2; Length 1508;
Best Local Similarity 44.3%; Pred. No. 1.5e-134;
Matches 670; Conservative 236; Mismatches 177; Indels 129; Gaps 23;

QY 2 ETKRRYKMKHKVKKHWVTVAVASGLITLGTTLGSSVSAETEQQTSKVVTKSEDDKAAS 61
Db 6 ERNVRKLYKSKSWIGGLIILSTIMLSMTATSONVADSTNTVTDKSVTVSNNTNQ 65
QY 62 ESSQTDAPKTKAQTEQTAQSOAN-----VADTSTSITKETPSQNIITTOANSDDK- 112
Db 66 HDTVVD-KQTIPIVKNDQTTQIAANATAQAEVKASDTTDTQKASTANTNKEDSIDNL 124
QY 113 -----TVNTKSEEAQTSSEERTKQSEEAQTTASSQALT-----OAKAE-LT 152
Db 125 TKQLPAVPTANQTKGYLEKQKGVYVTSNDTLKGLITVDNHNKQYFDNNGVQAKGQFVT 184
QY 153 KQKQRT-----AAQENKNPV--DLAAIPN----- 173
Db 185 DNSKTYLLDPNSGNVATGQIGSQTLAFNDNGEQVFADFYTAPDGKTYFFDKGQATIG 244
QY 174 VKQIDGKYYIGSDGQPKNFALTNNKVLVYDKNTGALTDTSQYQFKQGLTKLNNDYTP 233
Db 245 LKAINGHNYFDSLQGLKKFTGVIDQVRYFQDSGQEVSTDSQ1KEGLTSQTTDYTA 304
QY 234 HNQIVNFENTSLETTIDNYVTADSWYRPKDILKNGKTTWTASSSDLRPLMSWPDQTOI 293
Db 305 HNAVHSTDSADPNFNGVLTASSWYRPKDVLRNGHWEATTANFRPIVSVVMPKQTOV 364
QY 294 AYLNNMQGL-GTGENVTADSSQSLNLAATQVQVKIETKISQTOQTQWLRLDIINSFVK 352

QY	1122	DAKGNWYFDNNHGMVYGLQQLNGEVOYFLSNGVLRRESFLENADGSK-NYFCHGLNRYS	1180	QY	430	NPVQAEQLNWLHLYMNYGSIIVANDPEANFDCGVRVDAVD---NVNADLLQIASDYLKAHY	486
DB	1078	EISGSWYFSDSGKMATKTKICNDYLFPPNGKQLKEGV--YDGKAYYYDDNGRTWT	1135	DB	404	N-----NTYTFDSNGVM-----QTDWITIDSKYFVSNG--VMQTGIYISGY	447
QY	1181	N-GYYSF--DNDGKRYFDASGVMAVGLKTINGTOYFDODGVGVKGAWITSDGKKRYF	1237	QY	487	GVDKSEKNAIHLSLILEAWSNDNDPOVKNQDTKGAQLPIDNKLRLSLLYALTRPLEKASNK	546
DB	1136	NKGVEFRVDGQDKRYFNGDGTIAIGLSLDRNRLDYDAYGYQVKGQTVT-INGKSYTF	1194	DB	448	YGFANDGKLLTGLQVI-----NGNSYYFDTNGIRL-----VSRWITIDGKY	489
QY	1238	DDGSGNMVNRFAENDK-----NGDWYFLNSDGLALGVQVQTINGKTYFFGQDGRQ	1286	QY	547	NEIRSGLEPFIITNSLNNRSAGK-----NNSRMANYIFIRAHDE--VQTV	590
DB	1195	DADQGLVDTDNANPAPQAGWKLGDQWGY-RKDGQLLTGEQITIDQKVFQDNGVQ	1253	DB	490	YFNQDG---ILTDNWIN--YDGKYYFIISGVKQTLQNLID--GNYYVF--DSSGIMQTG	539
QY	1287	IRGKIITD-NGKLKYLFLANSBELARNIFATDSQNNWYFSGDVAVTGSQTIAGKLYPA	1345	QY	591	IAKIITKAQINPKTDGLTFLDE--LKQAFKI-----YNED--MRQAKKYTQSNITP	638
DB	1254	VKGCTATDASGLVRFYDRDQGHQVGKWTSTSDNNWYVNESQVLTLGLQTDGQTVYFD	1313	DB	540	LOKI-----DGKTYFYGNGIRQIGWITYQNNKYFNSDGSMTDLKYSYSTSY	590
QY	1346	SDGQVKGFSV-TYNGKVHYHADSGELQVNRFEADKQGNWYFLDSNGEA	1394	QY	639	AYALMLSNKDSITRLYYGDMYSDDG-----OYMATKSPYYDA--IDTL---LKARIKY	686
DB	1314	DKGIOAKGRAVMDENGLRYFDADSGNMLRDRWK-NVDGNWYFNRNGLA	1362	DB	591	NTHYO-----YYG--FNDGKLLTGLQTIKNTYFDSNGISQMGWNIDGKDFY	638
RESULT 13							
B97033							
uncharacterized protein, related to enterotoxins of other Clostridiales [imported] - Cl							
C:Species: Clostridium acetobutylicum							
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004							
C:Accession: B97033							
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,							
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.							
J. Bacteriol. 183, 4823-4838, 2001							
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo							
A:Reference number: A96900; MUID:21359325; PMID:21359325							
A:Accession: B97033							
A>Status: preliminary							
A:Molecule type: DNA							
A:Residues: 1-2817 <KUR>							
A:Cross-references: UNIPROT:Q97K42; UNIPARC:UPT00000CA0A0; GB:AE001437; PIDN:AAK79053.1;							
A:Experimental source: Clostridium acetobutylicum ATCC824							
C:Genetic:							
A:Gene: CAC1079							
Query Match 7.5%; Score 562; DB 2; Length 2817;							
Best Local Similarity 21.9%; Pred. No. 2e-16;							
Matches 351; Conservative 206; Mismatches 557; Indels 487; Gaps 77;							
QY	85	ANVADTSTITKETPSON-----ITQAN-SDDQVNTKSEEAQTSERTKOS	132	QY	1015	AFL--SELAAKYPSIFNRTQISNGKKIDPSEK-----ITAMK-----AKYFNGTNI	1058
DB	26	SELTQNSALTKRSSNNSFLNKQHVPTITSNVNGSNKNNLTK-VQNTASSMNPWT	84	DB	931	MLGYNINGKYYFNDNGVITQGWVTRSSKYYLDPGAAVTFQFQINGDKYFNSSGI	990
QY	133	EEAQTASSQALTOAKAELTKQRTAAQENKXNPVDLAAIPNVKQIDGKYYIYIGSDQPKK	192	QY	1059	LORGVGVULKD-----NASDKYFE-----LKGNOTYLPK-QMTNKEAS	1095
DB	85	NPQKATNSKILVNPKL----NQASSPNEGITEPKQASIPYTNVTDKNKTF-----K	132	DB	991	MOTGLVYVNPDPYFGFDDNGHILTMHSHINGIYIYFDSTGKAQKGFYTLGKTYFNTNMV	1050
QY	193	NFALTNNKVLVPDKNTGALTDTSQYQFKGLTKLANDYTPHNOIYVFNENTSLETTDNYV	252	QY	1096	TCFVNDGNGMTFYSTSG-----YQAKNSFV---QDAKGNWYFDNNHGMV	1137
DB	133	NES-SINNEAPIPKDTSKTKSTSSAQ-----TKGSDN--NNIPS-NNITSTNTSKNEN	181	DB	1051	TCFVANNMLLYFDNEGVMQGWINNSRYVFSATGASVTGFTDDGKNCYCDNSGAIY	1110
QY	253	TADSWRPRDILKNGKGTWTASSESDLRPLLMSWMPDKQTOIAYLNVMMQGLGTGENYTA	312	QY	1138	YGLQQLNGEVOYFLSNGVQLRESFLENADGSKNYFGLHGNRYSNGY-----YSPOND--	1189
DB	182	PSNT-----DI-----KTEPANAIPKDT-----PNNQSDSA-----LAKNKALSNRNLA	223	DB	1111	TDVVTINGSTYGFNTDGIMLT-----GWQIRYNRGYSYFNTYFNSDGT	1156
QY	313	DSQESLNLAAQTVQVKIETKISQTOQWLRIINSFVK-----TQPNWNSQ	360	QY	1190	-----SKRYFPDAS--GVMAVGLKTINGTOYFDQDGVQVKGAWITSDGKKRYFDD	1239
DB	224	DSSQTSKVITSSNNDAPKVNTTSTDKKASLNLDNSQDGVTKDGKKYYVNGVQKGFQSI	283	DB	1157	AKTGFTYLNKTYFYNFSPDSGRMLQGYQYINGNHYPAPDGTMTQTG-WIT--NGSSKYYLD	1213
QY	361	TESDTSAGEKDHLOGGALLYS-----NSDKTAYA-----	389	QY	1240	GSGNMVAV-----NRFANDKNG-----DWYFLNSDGLALGVQTINGKTYF	1280
DB	284	NKSIYFNDGSMQGTWLNKNSNYFSDASGVMLTGLQNINGTYFQFNDGKLLTGLQAI	343	DB	1214	PGSAVTVGLQTINGNKYCFDSNGILOHNGIFIGTYNTYGSDDNNGIMLTGLQLINGLYCF	1273
QY	390	NSDYRLNRTPTTSQTG-----KPKYFEDNS-----SGGYDFLLANDIDNS	429	QY	1281	GQDGKQIKGKIITDNGKLYFLANSBELARNIPATDSQNNWYFSGD-----	1327
DB	344	NNWYFNDNGVMQGTGWITCNDKSKYYFDNNGVMQGLVHNNKYFGNDGKLLTGLQNI	403	DB	1274	NSDG-SVKTGLVTLGKTYFDSYSVSGFQNI-----NNNTYFGNDGTMTGWNVGY	1327

Db 1391 DFGSDIDNKDRYIFLTCELDKSLIEINLVAKSYSLLSG-----DKNY 1436
QY 838 AASNKANATGQVYESSALDSQLI---YEGFSNFPDF--VTKDSYV---NKKIAQNVQL 889
Db 1437 LISNLNTIEKI--NTLGLDSKNIAYNYTDESNNKYFGAISKTSQKSIIHYKDSKNILE 1494
QY 890 FKSWGVTSPEMARQYVSSDGSFL-DSI--IQNGYAFEDRYDLA-----MSKN----- 934
Db 1495 FYNDSTLEFN-SKDFIAEDINVFMDKDINTITGKYVDNNTDKSIDFSISLVSKNQKVN 1553
QY 935 -----NKYGSQODMINAVKALHSG-----IQVIADW-----VPODIYNLPGK 972
Db 1554 GLYLNESVYSSYLDLVKNSDGHNTSNFNLFLDNISFWKLFCFENINVIDKYFLVGK 1613
QY 973 -----EVTATRVN---DYGEYRKDSEIKNTLYAANTKS-----NGKD----- 1007
Db 1614 TNLGYVEFICDNKNKNDIYGEW-KTSSSKSTIFSGNRNVVVEPIVNPDTGEDISTSLD 1672
QY 1008 --YQAKYG-----GAFSELAAKYPSIPNRTQISNGKK 1038
Db 1673 FSYEPLYGIDRYINKVLIAPDLYTSLININTNYSNEYYPEIIVLNPNTFHKKVWIN--- 1729
QY 1039 IDPSEKITAWK-----AKYFNGTN-----ILGRG 1062
Db 1730 LDSSEFYKWSGTEGDFILVYLEESNKKILQIRIKGILSNQTSFNKMSIDFKDIKLS 1789
QY 1063 VGYVL-----KDNASDK---YFELKGNQTYLPKQMTNKEASTGFVNDGNGMTFYS--- 1109
Db 1790 LGYIMSNFKSFENSELDRDLGFKIIDNKTYVDE--DSKLVKGLININNSLFYFDPPIE 1847
QY 1110 ---TSGYQAKNSFVQDAKGNWYFDF--NNGHWYGLQQLNGEVOYFLSNGVQLRESFLENA 1165
Db 1848 FNLVTGQMTIN-----GKKYFEDINTGAALTSYKIIINGKHFFYFNNDGV-MQLGVFKGP 1899
QY 1166 DGSKNYFGLHGNYSN-----GYYSFNDNSK-----YFDA 1197
Db 1900 DGFE-YFAPANTQNNIEGQAIYVQSKFLTLNGKKYTFDNNSKAVTGWRILINNEKYFNP 1958
QY 1198 -SGVMAVGLKTINGNTQYFDQDGQYQVKGAWITGSDGKGRYFDDGSGNMVNRFPANDKNGD 1256
Db 1959 NNAIAAAGLVQIDNNKYNFNPDTAIIISKGWQT-VNGSRYFYDDTDA-IAPNGY-KTIDGK 2015
QY 1257 WYILNSDGIALGV-----QTINGKTYFFQDGK 1285
Db 2016 HFYFDSDCVVKIGVFTSTNGFEYFAPANTYNNNIEGQAIYVQSKFLTLNGKKYTFDNNSK 2075
QY 1286 QIKGKIITDNGKLYFLANSSELARNIFATDSQNNWYFGSD-GVAVTGSQTIAGKKLYF 1344
Db 2076 AVTGWQTIDSKKY-YFNTNTAEAAATGWOTIDGKK--YFNTNTAEAAATGWOTIDGKKYF 2132
QY 1345 ASDGKQVKGSVFTYNGKVHYHADSGELQV-----NRPE-----ADKD----- 1382
Db 2133 NTNTAIASTGYTIINGK-HFYFNTDGIQIGVFKPGNPFYFAPANTDANNIEGQAILYQ 2191
QY 1383 -----GNWYILDSNGEALTSQRINDQRVFF 1408
Db 2192 NEFLTNGKKYFGSDSKAVTGWRIINNNKYYF 2224

RESULT 15
A37184
glucan-binding protein - Streptococcus mutans
C;Species: Streptococcus mutans
C;Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 09-Jul-2004
C;Accession: A37184
R;Banas, J.A.; Russell, R.R.B.; Ferretti, J.J.
Infect. Immun. 58, 667-673, 1990
A;Title: Sequence analysis of the gene for the glucan-binding protein of Streptococcus m
A;Reference number: A37184; PMID:90170123; PMID:2307516
A;Accession: A37184
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-563 <BAN>

A;Cross-references: UNIPROT:Q54447; UNIPARC:UPI00000B8CB5; GB:M30945; NID:g153637; PIDN
F;169-188/Domain: cpl repeat homology <CP1>
F;264-283/Domain: cpl repeat homology <CP2>
F;349-368/Domain: cpl repeat homology <CP3>
F;504-523/Domain: cpl repeat homology <CP4>
F;525-548/Domain: cpl repeat homology <CP5>
Query Match 5.8%; Score 435, DB 2; Length 563;
Best Local Similarity 26.8%; Pred. No. 6.8e-12;
Matches 137; Conservative 72; Mismatches 180; Indels 122; Gaps 19;
QY 989 DSEIK-NTLYAANTKSNKDYQAKYGGAFLSELAAKYPSIENRTQISNGKKIDPSPKITA 1047
Db 70 DSVENNSANAVKSBETAEEAVSDGR-----ASQTEAVTNTQ-----NSEEHHHPAKATA 121
QY 1048 WKAKYFNGTNIILGRGVGYVLKDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDGNGMTF 1107
Db 122 VSGE-----AQSVQNAPESEAAQ-----QETAKTEPATAEENDAAPT- 159
QY 1108 YSTSGVQAKNSFVQDAKGNWYFDFNNGHWYGLQQLNGEVOYFLSNGVQLR-ESFLENAD 1166
Db 160 -----NSPFK-KDGEWYIYKADQLATGQIIDGRKQLYFNQDGSQVKGIEIHVETGD 209
QY 1167 GSKNYFGLHGNYS-----NGYYSFNDNS-----KWRYEDA 1197
Db 210 QIIYHPVFIISDPSVLVKNKIYFDPDSELKWDKRFVYSSYADPLHYENIKHEGWYFLGE 269
QY 1198 SGVMAVGLKTINGNTQYFDQDGQYQVKGAWITGSD-----GKCRYFDDGSGNMVNR 1248
Db 270 DGKAAIGWITIGKKYFDTNGVQVKGKLISTDGNYNLISQKYGKKSFLDPDTGEAWTNR 329
QY 1249 FANDK-----NGDWYILNSDGIALGVQTINGKTYF-GQDGKQIKGIITDNG 1296
Db 330 FVNAKYTFYFNFAGYVSTTDFWYMGADGIGVTDWQKIDGMDYTFPEPSSGIVQKGDIAERDG 389
QY 1297 KLYFLANSSELARNIFATDSQ-----NWYFGSDGVA---VTGSQTIAGKKLYF 1344
Db 390 KVIYLDSDSGVVKRFGTTPAERISTVEARFPKTYIFGADGSRKDLTGWQIIDGKTYF 449
QY 1345 ASDGK-QVKGSVFTYNGKV-----HYHADSGELQVNRF--EADKGNWYILDS 1390
Db 450 KDDHSIKAKSEYSQIGSVDPDGPABIDGDGYFFDQGGQVFNTRFVRKYDYSNIWYIYGS 509
QY 1391 NGEALTSQRINDQRVFFTRE-----GKQVKG 1417
Db 510 DGRVSGWQTIIDGRYFSDQDEKTKGRQIKG 540

Search completed: February 11, 2006, 19:42:03
Job time : 38.5158 sec

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 18:57:35 ; Search time 192.956 Seconds
(without alignments)
5228.676 Million cell updates/sec
Title: US-10-797-821-36
Perfect score: 7462
Sequence: 1 METKRYKMHKVKHWTVA.....EGKQKGDVAYDERRLLVYR 1430
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	7412	99.3	1462	1 GTFD_STRMU	P49331 streptococc
2	4550.5	61.0	1577	2 Q54178 STRGN	Q54178 streptococc
3	4445.5	59.6	1575	2 Q9LCH3 STOR	Q9LCH3 streptococc
4	3860	51.7	1476	1 GTFB_STRMU	P08987 streptococc
5	3941.5	51.5	1454	2 Q69A94 LEUME	Q69A94 leuconostoc
6	3788.5	50.8	1577	2 Q55265 STRSL	Q55265 streptococc
7	3743	50.2	1455	1 GTFC_STRMU	P13470 streptococc
8	3713.5	49.8	1449	2 Q68542 STRSL	Q68542 streptococc
9	3713.5	49.8	1449	2 Q55264 STRSL	Q55264 streptococc
10	3661	49.1	1512	2 Q9WXJ5 STRE	Q9WXJ5 streptococc
11	3639	48.8	1506	2 Q56CX8 STRE	Q56CX8 streptococc
12	3615	48.4	1554	2 Q8KZL5 STRE	Q8KZL5 streptococc
13	3587.5	48.1	1597	1 GTFL_STRDO	P11001 streptococc
14	3575	47.9	1590	2 Q59983 9STRE	Q59983 streptococc
15	3544	47.5	1590	2 Q55263 9STRE	Q55263 streptococc
16	3542	47.5	1592	1 GTF2_STRDO	P27470 streptococc
17	3502.5	46.9	1518	2 Q00600 STRSL	Q00600 streptococc
18	3369.5	45.2	1561	2 Q58BM8 9LACO	Q58BM8 lactobacill
19	3334.5	44.7	2835	2 Q8G9Q2 LEUME	Q8G9Q2 leuconostoc
20	3312	44.4	1599	2 Q00599 STRSL	Q00599 streptococc
21	3252.5	43.6	1508	2 Q52224 LEUME	Q52224 leuconostoc
22	3251.5	43.6	1290	2 Q48756 LEUME	Q48756 leuconostoc
23	3245.5	43.5	1508	2 Q9EZH5 LEUME	Q9EZH5 leuconostoc
24	3227	43.2	1477	2 Q9L466 LEUME	Q9L466 leuconostoc
25	3186.5	42.7	1527	2 Q8KRE1 LEUME	Q8KRE1 leuconostoc
26	3174	42.5	1330	2 Q84CN4 LEUME	Q84CN4 leuconostoc
27	3174	42.5	1365	1 GTFS_STRDO	P29336 streptococc
28	3172.5	42.5	1522	2 Q6TXA4 LEUME	Q6TXA4 leuconostoc
29	3156.5	42.3	1527	2 Q9ZXV4 LEUME	Q9ZXV4 leuconostoc
30	3105.5	41.6	1338	2 Q9WXJ4 9STRE	Q9WXJ4 streptococc
31	3067	41.1	1595	2 Q58BM3 LACSK	Q58BM3 lactobacill

32	2723.5	36.5	1463	2 Q58BM6 LACFE	Q58BM6 lactobacill
33	2699.5	36.2	2057	2 Q9RE05 LEUME	Q9RE05 leuconostoc
34	2642	35.4	1016	2 Q9LCJ7 LEUME	Q9LCJ7 leuconostoc
35	2460.5	33.0	1772	2 Q58N3 LACRE	Q58N3 lactobacill
36	2447	32.8	1781	2 Q58BL9 LACRE	Q58BL9 lactobacill
37	2408.5	32.3	1781	2 Q4JCS4 LACRE	Q4JCS4 lactobacill
38	2405	32.2	1772	2 Q58N0 LACRE	Q58N0 lactobacill
39	2404.5	32.2	1781	2 Q4JLC7 LACRE	Q4JLC7 lactobacill
40	1552.5	20.8	591	2 Q8VUH3 STRMU	Q8VUH3 streptococc
41	1344.5	18.0	1619	2 Q58M0 LACRE	Q58M0 lactobacill
42	1330.5	17.8	1231	2 Q58N1 LACRE	Q58N1 lactobacill
43	1205	16.1	522	2 Q8VV10 STRSA	Q8VV10 streptococc
44	823	11.0	374	2 Q6ZXI9 9LACO	Q6ZXI9 lactobacill
45	562	7.5	2817	2 Q97K42 CLOAB	Q97K42 clostridium

ALIGNMENTS

RESULT 1

GTFD_STRMU STANDARD; PRT; 1462 AA.

ID GTFD_STRMU STANDARD; PRT; 1462 AA.

AC P49331; Q69383; Q69386; Q69389; Q69392; Q69398;

DT 01-FEB-1996 (Rel. 33, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)

DE (Sucrose 6-glucosyltransferase).

OS Name:gtfd; OrderedLocNames=SMU.910;

GN Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=GS-5;

RC MEDLINE=91100958; PubMed=2148600;

RA Honda O., Kato C., Kuramitsu H.K.;

RA "Nucleotide sequence of the Streptococcus mutans gtfd gene encoding the glucosyltransferase-S enzyme";

RT J. Gen. Microbiol. 136:2099-2105(1990).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f, MT4467 / Serotype e, and MT4148 / Serotype c;

RC MEDLINE=98231643; PubMed=9570124;

RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;

RA "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans";

RL FEMS Microbiol. Lett. 161:331-336(1998).

RN [3]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=UAI59 / ATCC 700610 / Serotype c;

RC MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;

RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P., Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A., Ferretti J.J.;

RA "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental pathogen";

RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.

CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-fructose + (1,6-alpha-D-glucosyl)(n+1).

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

CC CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC CC -|- SIMILARITY: Contains 6 cell wall binding repeats.
CC CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC EMBL; M29296; AAA26895.1; -; Genomic DNA.
DR EMBL; D88653; BAA26103.1; -; Genomic DNA.
DR EMBL; D88656; BAA26107.1; -; Genomic DNA.
DR EMBL; D88659; BAA26111.1; -; Genomic DNA.
DR EMBL; D88662; BAA26115.1; -; Genomic DNA.
DR EMBL; D89979; BAA26121.1; -; Genomic DNA.
DR EMBL; A8014932; AAN58619.1; -; Genomic DNA.
DR HSSP; P06653; LGVM.
DR InterPro; IPR002479; Cell wall bd put.
DR Pfam; PF003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
KW Transferase.
FT SIGNAL ? Potential.
FT CHAIN ? Glucosyltransferase-S.
FT REPEAT 1232 1295 1.
FT REPEAT 1296 1359 2.
FT REPEAT 1360 1423 3.
FT REGION 1232 1423 3 X 63 AA approximate tandem repeats.
FT VARIANT 10 10 Y -> H (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain WT8148).
FT VARIANT 19 19 I -> V (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain WT8148).
FT VARIANT 58 58 K -> E (in strain MT4239 and strain
FT VARIANT 68 68 A -> S (in strain MT4251 and strain
FT MT4245).
FT VARIANT 81 81 A -> T (in strain MT4251 and strain
FT MT8148).
FT VARIANT 113 113 T -> I (in strain MT4239 and strain
FT MT4245).
FT VARIANT 122 122 A -> V (in strain MT4239, strain MT4245
FT and strain WT8148).
FT VARIANT 132 132 A -> S (in strain GS-5 and strain
FT MT4467).
FT VARIANT 135 135 A -> V (in strain MT4245).
FT VARIANT 137 137 A -> T (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain WT8148).
FT VARIANT 202 202 V -> L (in strain MT4239).
FT VARIANT 255 255 D -> N (in strain WT8148).
FT VARIANT 275 275 E -> D (in strain MT4239, strain MT4245
FT and strain MT4251).
FT VARIANT 288 288 D -> N (in strain MT4239, strain MT4245
FT and strain MT4251).
FT VARIANT 301 301 Q -> H (in strain MT4245).
FT VARIANT 313 313 D -> N (in strain MT4239 and strain
FT MT4251).
FT VARIANT 317 317 E -> K (in strain MT4239).
FT VARIANT 328 328 V -> P (in strain MT4239).
FT VARIANT 350 350 F -> L (in strain MT4239, strain MT4251
FT and strain MT4467).
FT VARIANT 628 633 KKRYTQ -> EKEYTL (in strain MT4251).
FT VARIANT 688 688 A -> S (in strain MT4239).
FT VARIANT 726 732 TDGSEA -> ADKGND (in strain MT4251).
FT VARIANT 726 730 TDGGS -> ADKEN (in strain MT4239 and
FT strain MT4245).
FT VARIANT 762 762 T -> A (in strain GS-5, strain MT4239,
FT strain MT4245, strain MT4251, strain
FT MT4467 and strain WT8148).
FT VARIANT 964 964 D -> Y (in strain MT4251).
FT VARIANT 1019 1019 E -> K (in strain MT4245 and strain

FT VARIANT 1059 1060 MT4251).
FT VARIANT 1060 1060 LG -> IR (in strain MT4251).
FT VARIANT 1080 1080 G -> R (in strain MT4245).
FT VARIANT 1142 1142 H -> Q (in strain GS-5).
FT VARIANT 1198 1198 S -> N (in strain MT4239).
FT VARIANT 1220 1220 Y -> C (in strain MT4251 and strain
FT MT4467).
FT VARIANT 1280 1280 F -> L (in strain MT4467).
FT VARIANT 1282 1282 Q -> P (in strain MT4245).
FT VARIANT 1290 1290 K -> T (in strain MT4245).
FT VARIANT 1311 1311 N -> D (in strain MT4245).
FT VARIANT 1403 1403 G -> D (in strain GS-5 and strain
FT MT4467).
FT VARIANT 1425 1425 G -> R (in strain GS-5).
FT VARIANT 1449 1449 R -> K (in strain MT4467).
FT CONFLICT 1428 1462 RYDKNSGMVYKVVLTANGRRIGIDRWGIARY -> VY
FT R (in Ref. 1).
SQ SEQUENCE 1462 AA; 163388 MW; CE4A279C4D708645 CRC64;
Query Match 99.3%; Score 7412; DB 1; Length 1462;
Best Local Similarity 99.4%; Pred.No. 5.4e-304;
Matches 1420; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 1 METKRYKMHKVKHVVAVASGLITLGTTLGSSVSAETEQOTSDKVVTKSEDDKAA 60
DB 1 METKRYKMYKVKHVVAVASGLITLGTTLGSSVSAETEQOTSDKVVTKSEDDKAA 60
QY 61 SESSQTDAPKTKAQTEQTAQSQANVADTSTSIKTPSQNITTOANSDDKTVNTKSE 120
DB 61 SESSQTDAPKTKAQTEQTAQSQANVADTSTSIKTPSQNITTOANSDDKTVNTKSE 120
QY 121 EAQTSERTKQSEEAQTASSQALQAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
DB 121 EAQTSERTKQSEEAQTASSQALQAKAELTKQRTAAQENKPNVDLAAIPNVKQIDGK 180
QY 181 YYYIGSDGQPKNFALTNNKVLVFDKNTGALTDTSQYQKGLTKLNNDYTHPHNOIVNF 240
DB 181 YYYIGSDGQPKNFALTNNKVLVFDKNTGALTDTSQYQKGLTKLNNDYTHPHNOIVNF 240
QY 241 ENTSLETIDNVYTADSWYRPKDIKNGKWTWASSSDLRPLLMWMPDKQTOIAYLNYMN 300
DB 241 ENTSLETIDNVYTADSWYRPKDIKNGKWTWASSSDLRPLLMWMPDKQTOIAYLNYMN 300
QY 301 QQGLGTGENYTADSSQESLNLAATQVQVKIETKISQTOQTQWLRLDIINSFVKTPQNNWQ 360
DB 301 QQGLGTGENYTADSSQESLNLAATQVQVKIETKISQTOQTQWLRLDIINSFVKTPQNNWQ 360
QY 361 TESDTASAGEKHQGGALLYSNSDKTAYANSYRLLNRTPTTSQTKPKYFEDNSSGGYDF 420
DB 361 TESDTASAGEKHQGGALLYSNSDKTAYANSYRLLNRTPTTSQTKPKYFEDNSSGGYDF 420
QY 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDAVDVNNADLLQIASD 480
DB 421 LLANDIDNSNPVQAEQLNWLHLYLMNYGSIIVANDPEANFDGVRVDAVDVNNADLLQIASD 480
QY 481 YLKAHYGVDKSEKNAIHLHSIILEASDNDPQNKOTKGALPIDNKLRLSLLYALTRPLE 540
DB 481 YLKAHYGVDKSEKNAIHLHSIILEASDNDPQNKOTKGALPIDNKLRLSLLYALTRPLE 540
QY 541 KDAANKNEIRSGLEPVITNSLNNSAEGKNSERMANYIFIRAHDSVQVTVIAKIIKAQIN 600
DB 541 KDAANKNEIRSGLEPVITNSLNNSAEGKNSERMANYIFIRAHDSVQVTVIAKIIKAQIN 600
QY 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNITPTAYALMLSNKDSITRLYYGDMYS 660
DB 601 PKTDGLTFTLDELKQAFKIYNEDMRQAKKYYTQSNITPTAYALMLSNKDSITRLYYGDMYS 660
QY 661 DQGQWATKSPYDAIDTLLKARIKYAACQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
DB 661 DQGQWATKSPYDAIDTLLKARIKYAACQDMKITTYVEGDKSHMDWDYTGVLTSVRYGT 720
QY 721 GANEATDQGSSEATKTQGMVITSNNSPLKLNQNDKVIIVNMGAAHKNQEVRLPILLTTKQGL 780


```
Db 721 GANEATDQGSSEATKTOGMVITSNPCLKNQNDKIVNNGTAKHKNQYRPLLLTTKDGL 780
QY 781 TSYTSDAAAKSLYRKTNKDGELVFDASDIQGYLNPQVSGYLAWVPVPGASDNQDVRVAAS 840
Db 781 TSYTSDAAAKSLYRKTNKDGELVFDASDIQGYLNPQVSGYLAWVPVPGASDNQDVRVAAS 840
QY 841 NKANATGQVYESSALDSQLIYEGFNFQDFTKSDYTNKKIAQNVLFKSGVTSFEM 900
Db 841 NKANATGQVYESSALDSQLIYEGFNFQDFTKSDYTNKKIAQNVLFKSGVTSFEM 900
QY 901 APOVYSEDSGLDSIIQNGYAFEDRYDLAMSKNKKYGSQODMINAVKALHKSIGQVIAD 960
Db 901 APOVYSEDSGLDSIIQNGYAFEDRYDLAMSKNKKYGSQODMINAVKALHKSIGQVIAD 960
QY 961 WVPDQIYNLPFGKEVWATRVNDYGEYRKDSEIKNTLYAANTKSGDKYQAKYGCAGFSEL 1020
Db 961 WVPDQIYNLPFGKEVWATRVNDYGEYRKDSEIKNTLYAANTKSGDKYQAKYGCAGFSEL 1020
QY 1021 AAKYPSIFNRTOISNGKKIDPSEKITAWKAKYFNGTNILGRGVYVLDKNASDKYFELKG 1080
Db 1021 AAKYPSIFNRTOISNGKKIDPSEKITAWKAKYFNGTNILGRGVYVLDKNASDKYFELKG 1080
QY 1081 NOTYLPKQMTNKEASTGFVNDGCMTFYSTGYOAKNSFVQDAKGHWYFDDNNGHMYGL 1140
Db 1081 NOTYLPKQMTNKEASTGFVNDGCMTFYSTGYOAKNSFVQDAKGHWYFDDNNGHMYGL 1140
QY 1141 QOLNGEVQVFLSNGVOLRESFLENADGSKNYFCHLGNRSNGYYSFDNDSKRYPDASGV 1200
Db 1141 QOLNGEVQVFLSNGVOLRESFLENADGSKNYFCHLGNRSNGYYSFDNDSKRYPDASGV 1200
QY 1201 MAVGLKTINGNTQYFPDQGVQVKGAWITGSDGKKRYFDDGSGNMVNRFANDKNGDWYYL 1260
Db 1201 MAVGLKTINGNTQYFPDQGVQVKGAWITGSDGKKRYFDDGSGNMVNRFANDKNGDWYYL 1260
QY 1261 NSDGIALVGQVTNGKTYFYFGQDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
Db 1261 NSDGIALVGQVTNGKTYFYFGQDGKQIKGKIITDNGKLYFLANSSELARNIFATDSQNN 1320
QY 1321 WYFSGDGVAVTGSQTIAGKCLYFASDGQVKGSGFYTYNGKVHYHADSGELQVNRFEAD 1380
Db 1321 WYFSGDGVAVTGSQTIAGKCLYFASDGQVKGSGFYTYNGKVHYHADSGELQVNRFEAD 1380
QY 1381 KQGNWYLDNSGEALTGSRINDQRVFFTRGKQVKGVDVAYDERLLVY 1429
Db 1381 KQGNWYLDNSGEALTGSRINDQRVFFTRGKQVKGVDVAYDERLLVY 1429

RESULT 2
Q54178 STRGN
ID Q54178 STRGN PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfG;
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RX MEDLINE=96157064; PubMed=85861195;
RA Vickerman M.M.; Sulavik M.C.; Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
phase variants."
RL Dev. Biol. Stand. 85:309-314(1995).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RA Minick P.; Vickerman M.;
```

```
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U12643; AAC34483.1; -, Genomic DNA.
DR PIR; B41898; B41898.
DR HSP; P06653; 1HCX.
DR GO; CO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco hydro 70.
DR Pfam; PF01473; CW binding I; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;
```

```
Query Match 61.0%; Score 4550.5; DB 2; Length 1577;
Best Local Similarity 59.7%; Pred. No. 2.5e-183;
Matches 895; Conservative 178; Mismatches 314; Indels 111; Gaps 17;

QY 1 METRRYKHKVKKHVVAVASGLITLGTTLG-----SSV-SA 39
Db 2 MEKKVHKHKVKKHVVAVTSLALLVAPKALGLESGLVYADDANQVTVNVEQSAVQSK 61
QY 40 ETEQOTSDKVVTKSEDDKAASESQTDPKTKQAQTEQTOAQSOAN-VADTSTS----- 93
Db 62 DSEQTTSDK-----ATDSSQLEVKBOASSSKETYQASAATNPANEQTTOODKE 110
QY 94 -ITKETPSONITQANSDDKTVNTKSEEAQTSSEETKQSEEAQTASSOALTQAK---- 148
Db 111 VETSRDSRHELTQKTSDD-----SSEKSGSSQEPKQVADQAESTDTKTQAALQAKQDSR 163
QY 149 -----AELTKQRTQATAQENKPV-----DLA 169
Db 164 ANDOEBETTENVAKATVSDKIATPKKERLPEPAQRKESITEKMLAAQAQAAPVNTHEDD 223
QY 170 AIPNVKQIDGKYIYGSDGQPKKNFALTNNKVLVFDKNTGALTDTFSQYQFKGLTKLNN 229
Db 224 VLAHKITDGKYYVDDGTVKKNFVAVELNGKILYFPAETGALVDSNEYQFQOGTSLNN 283
QY 230 DYTPHNQIVNFENTSIETIDNTVTADSWYRPRKIDKNGKTTASSSDRLPLLSWMPDK 289
Db 284 EFTQKNFVGTDDKIDETVDGVLTAADSWYRPFILKDGKTWTASTETDLRPLLMWPPDK 343
QY 290 QTOIAYLNYMNOQGLGTGENYTADSSQESLNLAATVQVVKIETKISQTOQTOWLRDIINS 349
Db 344 RTQINLYNTMNOENLGIGA-FESKTEQVLITNAVQVQRKIBERI SKEGDTKWLRLMSA 402
QY 350 FVKTPQNNNSQTESDTSAEKDHLOGGALLYNSDKTAYANDYRLLNTPPTSQTGPKY 409
Db 403 FVKTPQNNIKTESBTTGTNKHLOGGALLYTNSDKTSHANSRYRLNTRPTTSQTGPKY 462
QY 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNLWLHLYMNYGSIIVANDPEANPDGVRVDAVDN 469
Db 463 FIDKSNNGYEFLLANDFNSNPVQAEQLNLWLHLYMNYGSIIVANDPEANPDGVRVDAVDN 522
QY 470 VNADLLQIASDYLKAHYGVDSKNAIHLSTILEAWSNDNDPQYNKDTKGAQLPIDNKJRL 529
Db 523 VNADLLQIASDYFKSRKYGVSESEALKHLSILEAWSNDNDPYNKDTKGAQLAIDNKJRL 582
QY 530 SLLIYALTRPLEKXDAKNKEIRSGLEPVTNSLNNSAEGKNSERMANIYIFIRAHSEVQT 589
Db 583 SLLYSFMRKL-----SIRSGVEPTITNSLNDSTENKNGERTANYIFVRAHSEVQT 634
QY 590 VIAKIQAQINPKTDGLTFLDELQKAPKIYNEDMRQAKKYTQSNIPYALMLSNKDS 649
Db 635 VIADIIRENINPNTDGLIFTMDELQKAPKIYNEDMRKADKKYQINPITPAHALMLSNKDS 694
QY 650 ITRLYGDMYSDGQYMATKSPYDAIDTLLKARIKYAAGQDMKITVY----EGDKSHM 705
Db 695 ITRVYVGLYTDGQYMEKSPYDAIDALLRARIKYVAGQDMKITVYGVPREADK--- 751
QY 706 DWDYTGVLTSVRYGTGANEATDQSEATKQGMVITSNNPSLKNQNDKVIWNMGAH 765
Db 752 -WSYNGILTSVRYGTGANEATDEGTAETRTQGMVITASNPNLKLNEWDLQVNMGAH 810
```



```
QY 898 FEMAPQVYSEDGSLDIIQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHSGIQV 957
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
942 FELPQYVSQDGTFLDIIQNGYAFEDRYDYMAMSKNNKYGSLKDLINLRALHVSNIQA 1001
QY 958 IADWVDDQIYNLPGKEVWATRVNDYGEVHKDSEIKNTLYAANTKNGKDYQAKYGAFL 1017
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1002 IADWVDDQIYNLPGKEVWATRVNNTGTVREGAEIKELKVANSKNTNETDFQKYGGAFL 1061
QY 1018 SELAAKPSIFNRTQISNGKKIDPSBKITAWKAKYFNGTNILGRGVYVLKDNASDKYFE 1077
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1062 DELKAKYPEIFERVOYLSNGQKMTTDEKITKWSAKYFNGTNILGRGAYVYLKOWSNDYLT 1121
QY 1078 LKGNQTYLPGKMTNKAESTGTVNDGNGMTFYSTSGQAOKNSFVQDAGKNGWYFDNNGHMV 1137
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1122 NNRNGEVLPLKQLNKNYSYTGFSVDANGTKFYSTSGYQAOKNSFTQDENGWYFDKRGYLV 1181
QY 1138 YGLQQLNGEVOYFLSNGVOLRESFLENADGSKNYFGLNRYNSYVSDNSKRYEYFA 1197
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1182 TGAHEIDGKHVPLKNGIQLRDSIREDENGNQYYDQTGAQLNRYITTDGQN-WRYFDA 1240
QY 1198 SGVMAYGLKTINGNTQYFDQDGYQVKGAMITGSDGKKRYFDDGSGNNVNRFA-NDKNGD 1256
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1241 KGVMARGLVKIGDQFFDENGYQVKGKIVSAKDGKLYRFDKDSGNVAVINRFAQGNPSD 1300
QY 1257 WYTLNSDGLALVGVQTINGNTYTFYFGDQKQIKIITDNGK-LKYFLANSGLARNIFAT 1315
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1301 WYTFGVFEAKLTGLQKIGQOTLYFDQDQKQVKGKIVTLDKSIYRFDANSSEMAGVKFAE 1360
QY 1316 DSQNNWYFGSDGVATGVSQTTAGKLYFASDGKQVKGSGFVYNGK-VHYHADSGLQV 1374
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1361 GARNWYFYFDTKGAVTGKQKQKQVLYFDQDQKQVKGKIVTLDKSIYRFDADSGEMAV 1420
QY 1375 NREADKQNGWYTLNSGEBALTSQRINDQRFVFTREGKQVKGDV 1419
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1421 GRFAEGAKNEWYFDQTKAVTGLQKIDKQTLFYDQDQKQVKGKI 1465

RESULT 4
GTFB STRMU
ID GTFB STRMU STANDARD; PRT: 1476 AA.
AC P08387; 069381; 069384; 069387; 069390; 069396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (sucrose 6-glucosyltransferase).
GN Name=gtfB; OrderedLocustNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX NCBI_TaxID=1309;
RN [1]_TaxID=1309;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RN [2] J. Bacteriol. 169:4263-4270(1987).
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
RX MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RN [3] FEMS Microbiol. Lett. 161:331-336(1998).
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501293;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
```

```
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -I- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -I- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -I- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -I- SIMILARITY: Contains 10 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC -----
CC EMBL; M17361; AAA88588.1; -; Genomic DNA.
CC EMBL; D88651; BAA26101.1; -; Genomic DNA.
CC EMBL; D88654; BAA26105.1; -; Genomic DNA.
CC EMBL; D88657; BAA26109.1; -; Genomic DNA.
CC EMBL; D88660; BAA26113.1; -; Genomic DNA.
CC EMBL; D89977; BAA26119.1; -; Genomic DNA.
CC EMBL; AE014940; AAN58705.1; -; Genomic DNA.
CC F1R; B33135; B33135.
CC HSSP; P06653; IH8G.
CC InterPro; IPR002479; Cell_wall_bd_put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_I; 4.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
KW Transferase.
RN [1] SIGNAL
RP CHAIN 35 1476 Potential.
FT REPEAT 1097 1130 Glucosyltransferase-I.
FT REPEAT 1161 1210 A repeat.
FT REPEAT 1225 1275 1.
FT REPEAT 1290 1340 2.
FT REPEAT 1355 1405 3.
FT REPEAT 1420 1470 4.
FT REPEAT 1420 1470 5.
FT REGION 35 1051 Catalytic (approximate).
FT REGION 1097 1476 Glucan-binding (approximate).
FT VARIANT 62 62 S X tandem repeats.
FT VARIANT 65 65 S -> T (in strain MT4239).
FT VARIANT 68 68 T -> I (in strain GS-5).
FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245,
FT strain MT4251, strain MT4467 and strain
FT MT8148).
FT VARIANT 78 78 Q -> P (in strain MT4251).
FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245,
FT strain MT4251, strain MT4467 and strain
FT MT8148).
FT VARIANT 89 89 S -> F (in strain MT4251).
FT VARIANT 168 168 K -> N (in strain MT4251).
FT VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and
FT strain MT8148).
FT VARIANT 399 399 N -> R (in strain MT4239).
FT VARIANT 474 474 I -> T (in strain MT4239).
FT VARIANT 512 512 K -> R (in strain MT8148).
FT VARIANT 519 519 P -> Y (in strain MT8148).
FT VARIANT 701 701 T -> I (in strain MT8148).
FT VARIANT 708 708 A -> V (in strain MT8148).
FT VARIANT 938 938 F -> L (in strain MT8148).
FT VARIANT 952 957 FGKPYE -> YGTPVA (in strain GS-5, strain
```

FT	VARIANT	963	964	MT4239 and strain MT4467).
FT	FT			SV -> NT (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	968	970	ADS -> VDG (in strain GS-5, strain MT4239 and strain MT4467).
FT	VARIANT	1086	1086	A -> T (in strain MT4239).
FT	VARIANT	1158	1158	S -> N (in strain MT4239).
FT	VARIANT	1163	1163	H -> Y (in strain MT4251).
FT	VARIANT	1168	1168	E -> K (in strain MT8148).
FT	VARIANT	1182	1182	Y -> C (in strain MT8148).
FT	VARIANT	1234	1234	A -> P (in strain MT4239).
FT	VARIANT	1263	1263	R -> H (in strain GS-5 and strain MT4467).
FT	VARIANT	1263	1263	R -> P (in strain MT8148).
FT	VARIANT	1264	1264	Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1272	1272	S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1329	1329	H -> Y (in strain GS-5 and strain MT4467).
FT	VARIANT	1394	1394	Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
FT	VARIANT	1402	1402	S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT4467).
FT	VARIANT	1459	1459	Y -> H (in strain MT4467).
FT	CONFLICT	570	570	R -> A (in Ref. 1).
FT	CONFLICT	800	817	ADQDVRVAASAPSTDGK -> LIKMFALRLARPHQOWA (in Ref. 1).
FT	CONFLICT	1310	1310	H -> L (in Ref. 1).
FT	CONFLICT	1478	1478	MM; 9C6E09F731B4CBCF CRC64;
FT	SEQUENCE	1478	1478	AA; 165847 MM; 9C6E09F731B4CBCF CRC64;
Query Match				
Best Local Similarity				
Matches 765; Conservative 233; Mismatches 391; Indels 50; Gaps 20;				
Qy	1	METKRYKMKVKKHWTVA	VASGLITLGT	TTLGSSVSAETEOQTSKVVTKSEDDKAA 60
Db	1	MDKKVRYKLRKVKRWTVS	VASAVMTL	-TTLGSLGVKADSNESK-----QISNDSNTS 54
Qy	61	SESSQDAPKTKQAQTEQ	TAQSQANVADT	STSIKTPPSONITTAQNSDDKTVTNTKS- 119
Db	55	VVTANESNVTEVTSKQEA	ASSQTN--	HTVTTTSSSTSVVN--PREVVSNPTVGETASN 111
Qy	120	-EEAQTSEERTKQSEEAQT	TASSQAL	TOAKAELTKQRTAAQENKNPVDLAAIPNVKQID 178
Db	112	GEKLQNTTTVDKTSEAA	NNISKQTEAD	TDVDDSNAA-----NLQILEKLPNVKEID 166
Qy	179	GKYYYIGSDGPKKNFAL	TVNNKVL	YFDKNTGALTDTS-QYQFQKGLTKLANNYPHNOI 237
Db	167	GKYYYDNNKGVRTNFTL	ADGKILHFD	-TCAVTDTSIDTVNKOIVTTRSNLYKKNQV 225
Qy	238	VNFENTSLETTDNYV	TADSWRPK	DILKNGKTTWASSESDLRPLLMWSWDPKQOIAVLN 297
Db	226	YDRSAQSFHVDHYLTAE	SWRPYKYL	KDGKTTWQSTEKDPRPLMTWMPSQETQROXYN 285
Qy	298	YMNQOGLGTGNYTAD	SSOESLN	AAQTVQVKIETKISOTOOTOWLRDIINSFVKTPQNW 357
Db	286	YMNQA-IGINKTYD	TSNQLQNTAA	TIQAKIEAKITTLKNTDMLRGTISAFVKTSQAW 344
Qy	358	NSQTESDTSAGEKDH	LOGGALLYNS	DK-TAYANSDYRLNRTPTTSQTK--PKYFEDNS 414
Db	345	NSDSEKPPD----	DHLQNGA	VLVDNEGKLTPYANSNRYILNRTPTNQTKGKDPRYADNT 400
Qy	415	SGGYDFLLANDI	DNSNPVQA	EQLNLWLYHLYANYGSIVANDEANPFGVRVDAVDNVNADL 474
Db	401	IGGYEFLLANDVD	NSNPVQA	EQLNLWFLHFNFGNIYANDPDANPDSIRVDAVDNVADL 460
Qy	475	LQIASHVYLKAH	GYVDKSEK	NAIHLHSILEANSDDNDPQYNKDTKGAQLPIDKRLLSLYA 534
Db	461	LQIAGDYLKAAG	KIHLKNDKA	ANDHLHSILEANSDDNDTPYLDHGDGMNIMNDKRLLSLIFS 520
Qy	535	LTRPLEKDA	SNKNEIR	GLEPVTITNSLNNRSAEGKNSERMANYIPIRAHDSEVQTVIAKI 594

521 LAKPLNQ-----RSGMPLITNSLVNRDTDDNAETAAPVSPVSFFIRAHDSVEQDLDRLDI 577
595 IKAQINPTDGLTFTLDELKQAFKIYNEDMRQAKCKYQTQSNIPTAYALMISNKDSITRLY 654
573 IKAIEINPVVGVYSTMEIEIKKAFIYNKDLLEATEKKYTHYNTALSAYALLTNKSKSVPRVY 632
655 YGMYSDGQYMATKSPYDAIDITLKKARIKYAAGGQDMKIITYVEGDGSHMDWDYTGVLTT 714
633 YGDMFTDDGGYMAHKTYINBAIETLLKARIKYSGGQAMENQQVGNSE-----IIT 683
715 SVRYGTGANEAATDQGEATKTCQGMVITTSNNPSILKLNNDKVIYNMGAHKNOBYRPLLL 774
684 SVRYGKALKATDGTDBRTTTSVAVIEGNPNPSRLKASDRVVVNMGAHKNOBYRPLLL 743
775 TTQDGLTSTYSDAAAKSLYKTKNDKGELVPDASDIQGYLNPQVSGYLVAVWPVGASDNQD 834
744 TTDNGIKAYHSDQEAAGLVRYTNDRGELIIFTAADIKGYANPQVSGYLVGVWVPVGAADQD 803
835 VRVAASKANATQGVYSSSALDSOLIIYEGFSNFDQFTVKDSDTYNKKIAQNVLFKSWG 894
804 VRVAASTAPSTDGSKVHQNAALDSRVNFEFSNFQAFATKKEEYTNVVIKNDVKDKFAEWG 863
895 VTSFENAPQVSVSEDSGFLDSIIQNGYAFEDRDYDLAMSKNKYKQSGQDMINAVKALKHSG 954
864 VTDFENAPQVSVSDGSLDSVILQNGYAFETDRDGLGSKPNKYGTADDLVKAIKALKHSG 923
955 IQVIADVPDQIYNLPKGEYVTVATRVNDYGEYRKDSIEKNTLYAANTKSKNGKQYQAKYGG 1014
924 IKWMADVPDMYAPPEKEVVVTRVDKFKCPVEGSOIKSVLYIYVADSKSGKQDQAKYGG 983
1015 AFISELAAKYPSIFNRTQISNGKKIDPSEKITAUKAKYFNGTNTILGRGVYVVLKDNASDK 1074
984 AFLEELQAKYPELFAKQIISTGVPMDPSPVKIKQWSAKYFNGTNTILGRGAGYVYLKDOATNT 1043
1075 YFELKGNQ--TYLPKQMTNKEASTGVFVNDGNGMTFTYSTGVQAKNSFVQDAKGNWYFDN 1132
1044 YFNISDNKEINFPLKTLNLDQGVGSYDGKGYVYYSTSGVQAKNTYFISEG-DKWYFDN 1102
1133 NGHWYGLQQLNGEVOYFNLNGVQLRSEFLNADGSKNYFEGHGNRSNGYSYFSDNDSKW 1192
1103 NGYVNTGAQSGINGWYVYFNLNGQLRDLILKNEDGTAYYVNDGRRYENGYYQF-MSGVW 1161
1193 RYFDASGMAVGLKTINGNTQYFDQDQYQVKGAWITSGDGKKRYFDDGSGNMAVNRFPAND 1252
1162 RHFN-NGEMSGLTVIDGQYQYFDEMGYQAKGKPVTTADGKIRYFDFKQSGNMYNRNFTEN 1220
1253 KNGDWYLLSDGIALVGVQTINGKTYTYFPGDGQKIKKIITDN--GKLKYFLANSGLAARN 1311
1221 BEGKWLVLGSDGAAVTSQTINGOHLVFRANGVQVKGFEVFTDYGRIYSYDSNGDQIURN 1280
1312 IFATDSQNNWYFYFGSDGVAVTGSGOTIAGKKLYFASDGKQVKGSPVT--YNGKVHYVHADSG 1370
1281 RFVRNAQGWYFEDNNGYAVTGARTINGOHLVFRANGVQVKGFEVFTDHBGRISYYDGNSSG 1340
1371 ELQVNRFEADKDGNNWYLDNNGEALTGSGRINDQRVFTTBREGKQVKGDVAVDERLLVY 1429
1341 DQIRNRFRVNAQGWYFEDNNGYAVTGARTINGOHLVFRANGVQVKGFEVFTDYGRIYSY 1399

RESULT 5
Q69A94 LEUME
ID Q69A94 LEUME PRELIMINARY; PRT; 1454 AA.
AC Q69A94;
AD Q69A94;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Dextranucrase (EC 2.4.1.5).
GN Name=dsrP;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NEOTIDE SEQUENCE.

RESULT 5

069A94 LELME

ID O69A94 LEUME PRELIMINARY; PRT; 1454 AA.

AC Q69A94:

DT 25-OCT-2004 (TREMBlrel. 28, Created)

DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)

DE Dextranucrase (EC 2.4.1.5).

GN Name=dsrP;

OS Leuconostoc mesenteroides.

OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.

NCBI_TaxID=1245;

PN 22

RP NUCLEOTIDE SEQUENCE:

RC STRAIN=IBT-PQ;
RT Fernandez-Vazquez J.L., Lopez-Munguia A., Olvera C.;
RL "Molecular characterization of a dextranucrase gene from Leuconostoc
mesenteroides IBT-PQ isolated from pulque.";
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY504865; AAS79426.1; -; Genomic DNA.
DR GO: GO:0047849; F:dextranucrase activity; IEA.
DR GO: GO:0016757; F:transferase activity; IEA.
DR GO: GO:0009250; P:glucan biochemistry; IEA.
DR InterPro: IPR003318; Glyco hydro 70.
DR Pfam: PF01473; CW binding 1; 1_70.
DR Pfam: PF02324; Glyco hydro 70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;

Query Match 51.5%; Score 3841.5; DB 2; Length 1454;
Best Local Similarity 53.1%; Pred. No. 1.8e-153;
Matches 768; Conservative 208; Mismatches 411; Indels 59; Gaps 21;

QY 6 RYKHVKYKHVVAVAVASGLITLITL-----GSSVSAETEQTSDKVVYQK 53
DB 11 RKQMYKSGK--MLVIAGSVIIIGVTSFIQQAQADVSKQGVVVVTTAVNQNSDATTTDK 67

QY 54 S--EDDKA--ASESSQTDAPKQAOQTEQTAQSOA--NVADTSTSTKTPSQNIITQAN 108
DB 68 SITTPKATTTADTSTND---KATTTADTSTNDKATTTADTSTN-DKATTTADISTNNK 122

QY 109 SDDKVTNTKSEBAQTSEBTKQSEBAQTASSQALTOAKAELTKQRTAAQENKPNVDL 168
DB 123 ATTTADTSTNNKATTTAD--TSDNNASATTSKDVSSLA-----QKSAIDNNSKTTDT 174

QY 169 AA-----IPNVKQIDKYYIGSDQPKKFNALTNNKVLVYFDKNTGALTDTTSOYQKOG 223
DB 175 AASFEASSKNLTKIDKTYTYDNGQIKNFATVDGKVLVYFDKNTGALTDTTSOYQKOG 234

QY 224 LTKLNNDYTPHQINQVFNFTSLETIDNYTADSWYRPKDKILKNGKWTWASSSDRPLLM 283
DB 235 LFSNNNYTEHNASVGTSSASTYNTVDGYLTADSWYRPKDKIFVNGQWESSKDDDLRPLLM 294

QY 284 SWPDKQTOIAYLVNMQ--QGLGTGENTYADSSQSLNLAQTVQVKIETKISQTOQTOW 342
DB 295 TWMPKATHVNLNAMYLDATETVYTSDDSDQALNKAQNIQVKEIKSSEVQOTOW 354

QY 343 LRDIIINSFYKQPNWNSQSDTSAGEKDHLOGGALLYNSDKTAYANSDYRLNKRTPS 402
DB 355 LKDDISKFDVDSQNNNIAESKGT----DHLGGALLYNSDKTPDANSDYRLNKRTPN 410

QY 403 QTKPKPYFDNDSGGYDFLLANDINDSNPNVQAEQNLNHLHYLYNYSIVANDPEANFDGV 462
DB 411 QTGTPLTYTDPQTGGYDFLLANDVDSNPVQAEQNLNHLHYLYNYSIVANDPEANFDGI 470

QY 463 RVDADVNNADLQIADSYLKAHYGVYDKSEKNAIHLSTLEAWSNDNDPOYNKDTKGAQLP 522
DB 471 RVDADVNDADLQIADSYLKAHYGVYDKSDAISNQHVSILEWSDNDNAEYVNDGNDQLS 530

QY 523 IDNKLRLSLYALTRPLEKDAKNKEIRSGLEPVTISLNNRSAGEKNERMANYIFIRA 582
DB 531 MDNKLRLSLKYSLTWPPVDHYGNK---RSGLEPVTISLNNRSAGEKNERMANYIFIRA 587

QY 583 HDSEVQTVIAIKIQAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYQTSNIPATYAL 642
DB 588 HDSEVQTVIAIKIQAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYQTSNIPATYAL 647

QY 643 MLSNKDSITRLYGYDMYSDGOYMATKSPYDAIDTLLKARIYAAGGQDMKITVYEGDK 702
DB 648 ILTNKDTVPVRYGYDMYSDGOYMATKSPYDAIDTLLKARIYAAGGQDMKITVYEGDK 707

QY 703 SHMDMDYTGVLTSVRYGTGANEATDQSEATKQGMVITSNPSSLKLNQNDKLVNMG 762
DB 708 SWAADSRYGLTSSVRYGTGANEATDQSEATKQGMVITSNPSSLKLNQNDKLVNMG 767

QY 763 AHKNOEYRPLLTTKDGLTSYTSDA--AAKSLYRKTNDKGLYDFDASDIQGYLNPQVSGYL 821

DB 768 AHKNOEYRPLLTTKDGLTSYTSDA--AAKSLYRKTNDKGLYDFDASDIQGYLNPQVSGYL 827
QY 822 AVWVPVGASDNQDVRVAASNKANATGOVYESSSALDSQLIYEGFSNFDQFVTKSDVTNK 881
DB 828 AVWVPVGASDQDQARTESSTATTDDQTLHNSNAALDSQVYFSFNSPQSTPTTEAEYANV 887
QY 882 KIAQNVQLFKSGMGTSPFEMAPQVYSSDEGSLDSIIQNGYAFEDRYDLAMSNNKYSSQ 941
DB 888 QIANNTDLYKSWGITNFEPPQVRSSTSDSLDSIIQNGYAFEDRYDLGENTPTKYGTVD 947
QY 942 DMINAVKALHKSIGIQAIVADWPDQIYNLPKEVWTATRVNDYGEYRKDSIKNLTLYAANT 1001
DB 948 QLRTAIALHATGAIKAMADWPDQIYNLTGKEVAVQVRVNNSGIYNQDSVINKLTYSQT 1007
QY 1002 KSNQKDYQAKYGGAPLSLAALKYPSIFNRTQISNGKKIDPSEKITTAKKAYENGNTILGR 1061
DB 1008 VGGG-EYQALYGEFDEIKLPSLFEKNQIISTGVPMASEKIKWSAKYFNGTNIQGR 1066
QY 1062 GVGYYVLKDNASDKYFEL---KGNQTYLPKQMTNKEASTGFVNDGNGMTFYSTSGYQAKNS 1118
DB 1067 GAYYVLKDWATNEYFKVSTSSNSVFLPKQLTNEESNTGFISTDGGWYTYSTSGYQAKDT 1126
QY 1119 FVQDAKGNYYFDNNGHMYGLQNLGEVQYFSLNSGVQLRESFLENADGSKNYFGLGNR 1178
DB 1127 FIQDDKSNYYFDKNGYTYGQTVNDNNYYFLPNGIELQDAILEDKSGDVYFENQYKQ 1186
QY 1179 YSNGYYSFNDKSWRYFDASGVMA-VGLAKTI---NGNTQYFDQDQYGVKGAMITGSDGK 1233
DB 1187 TIDGYFLMLANKT-WRYFDKNGVNNAGLTVTVDGQKHQYDPKNGIQVKGTSVKQADGK 1245
QY 1234 KRYFDGSGNMVNFANDKNGDWWYLLNSDGLALVGQVQTINGKTYFYFGQDGKQIKGKIIT 1293
DB 1246 LAYEDTDSCEWNTNFEKTDGTWYFGADGLAVTGAQTIISGKLPFDADGQIKGKEAT 1305
QY 1294 D-NGKLYKFLANSSELARNIFATDSNNWYFSGSDGAVTGSQTIAGKKLYPASDGKQVK 1352
DB 1306 DKKGKMHYYDADSGEMTTNRFEKLSGDSWYFNKNGVNTGAQVINGQHLFFESNGNQVK 1365
QY 1353 G-SFVYNGKHYHYHADSGELQVNFADKNGDWWYLLNSDGLALVGQVQTINGKTYFYFGQDGKQIKGKIIT 1411
DB 1366 GREYATDGMRYDADSGDMVTNRPERISDSGWAYFGANGVAVTGAQNINGQOOLYPDTK 1425
QY 1412 GKQVKG 1417
DB 1426 GHQVKG 1431

RESULT 6

Q5265 STRSL
ID Q5265 STRSL PRELIMINARY; PRT; 1577 AA.
AC Q5265
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Name=gtfM;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25375 possesses at least two genes
coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621 (1995).
DR EMBL: L35928; AAC41413.1; -; Genomic_DNA.
DR FIR; T30858; T30858.
DR HSP; P06653; IHB8.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0009250; P:glucan biosynthesis; IEA.

DR InterPro; IPR004829; Csurface_antigen.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR ProDom; PD153432; Csurface_antigen; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1577 Glucosyltransferase.
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB898A7D3A7BF3 CRC64;

Query Match 50.8%; Score 3788.5; DB 2; Length 1577;
Best Local Similarity 49.1%; Pred. No. 3.4e-151;
Matches 777; Conservative 208; Mismatches 412; Indels 185; Gaps 21;

QY 1 METKRYKMKHKVKKHWTVAVAS-----G 24
DB 1 MENKVPFLKHKVKKWVTIGVTLSNVALAGSLLAQKVEADETSAPNGDGLQQLSEDG 60
QY 25 LITLGT-----LGSVSAETEQQTSDDKVVTQKSEDDKAASESSQTDAP 69
DB 61 TASLVTTTTEQASQAQSVASAVATASVSHETSFOAATSQAVSQAQTSQVSAQEV 120
QY 70 KTKQAQT--EQTAQSQANVAOTSITKETPSQ----- 101
DB 121 VSSQTQSSQQTQTTTQVSSQQTSTQVAGQTSQVTEQARPRVLNAAPAIATRAA 180
QY 102 -----NITQAASSDDKVT-----NTKSEEAQTSBERTKQSEEAQTASSQAL 144
DB 181 DSTIRINANRNTTITAGTTPNTVITIGPNTPKENVTVTPNGTRPNVTIVQPNQN 240
QY 145 TQAKAELTKQRTAAQENKPNVDLAAI-PNVKQIDGKYIYIGSDGPKKQNFALTNNKVL 203
DB 241 KPQSPQSPQPNKVPQNPQSLDYKPVASNLKTDGKYV-ENGVVKKNAATLDCRLY 299
QY 204 YDKNTGALTDTSTQYQFQKGLTKLNDYTPHNOIYNFNTSLETINDYNTASWYRPKDI 263
DB 300 YFDE-TGAMVDQSKPIYRADAI PNNSIYAVYQAYDTSSKFEHLNFLTASWYRPKQI 358
QY 264 LKNGKWTWASSDLRPLLMWMPDKQTOIAYLNYNQOGLGTGENYATDSSQESLNLA 323
DB 359 LKDGKWTWASTKDYRPLMTWMPDKVTQVYLYNYSQQFG-NKTYTTDMSYDLAAA 417
QY 324 QTQVQKIETKISQTOQWLRIINSFVKTPQPNWSQTESDTSAGSKDHLQGGALLYSNS 383
DB 418 ETVQGIETIERIGREGNTTTLRLQMSDFIKTPQGNWSEEDNLLVG-KDHLQGGALTFLNN 476
QY 384 DKTAYANSYRLNRTPTQTKPKPKYFEDNSSGGYDFLLANDIDNSNPVQABOLNWLHY 443
DB 477 SATSHANSDFRLMNRPTTQGTGRKYHIDRSNGGYELLANDIDNSNPVQABOLNWLHY 536
QY 444 LMNYGSIIVANDPEANFDGVRVADVADVNADLLQIASDYLKAHYGVDKSEKNAIHLISILE 503
DB 537 IMNIGSILGNDSANFDGVRIDAVNDVADLLQIASDYKPKYRVADNEANAHLISILE 596
QY 504 AMSDNDPQYNKDTKGALPTDNKRLSLIYALTRPLEKDAASNNKNEIRSGLEPVITSLNN 563
DB 597 AMSYNDHQYNKDTKGALSLIDNRLTLLTTLPLR-----KSNYRGSLSERVITSLNN 648
QY 564 RSABGNKSERMANYIIFRAHDSVQVVIKIIKAQINPKTDGLTFTLDELKQAFKYNED 623
DB 649 RSSEQKHTPRANDYIFVRAHDSVQAVLANIISKQINPKTDGFTFTMDLQKAFENYAD 708
QY 624 MRQAKKYQTSNPPTAVALMLSNKDSITRLYGDYSDGOYNATKSPYVDAJDTLLKAR 683
DB 709 IAKADKKYTOYNIIPAAVATMLNKSITRYYGDLFTDDQYNAEKSPYNAIDALLRAR 768
QY 684 IKYAAGQDMKITTYVEGDKSHMDWDYTVGLTSVRYGTGANEATDQGEATKTQGMVITS 743
DB 769 IKYVAGQDMKITTYVEGDKSHMDWDYTVGLTSVRYGTGANEATDQGEATKTQGMVITS 743
QY 744 NNPSLKLNDQKIVVNMGAHKQYRPLLLTTTKDGLTSYTSDA-AAKSLYRKTNDKGL 802

DB 820 NRPDKMLGDLNVLVNMGAHKQYRPLLLSKSTGLATYTKDSDVPAGLVRYTDNQNL 879
QY 803 VFDASDIQGYLNPOVSGYLAVWVPVGSADNQVRVAASKANATQVYESSALDSQLY 862
DB 880 TFDADDIAGHSTVEVSGYLAVWVPVGSASENQDARTKASSTKKE-QVPFESSALDSQVY 938
QY 863 EGFNSFPQDFTVKDSDYTNKKTIAQNVQLFKSMGVTSPFEMAPQVYSSSEDGFLDSIIQNGYA 922
DB 939 EGFNSFPQDFTVKDSDYTNKKTIAQNVQLFKSMGVTSPFEMAPQVYSSSEDGFLDSIIQNGYA 998
QY 923 FEDRYDLAMSKNKYKQSQODMINAVKALHKSIGIQTADWVPDQIYNLPKQEVVTAIRVND 982
DB 999 FEDRYDLAMSKNKYKQSLKDLMDALRALHAEGISALADWVPDQIYNLPKQEVVTAIRVND 1058
QY 983 YGEYRKDSEIKWTLVAANTKSKNGDYQAKYGGAPLSELAAPSPINRTQISNGKIDPS 1042
DB 1059 YGTPRPAEINSLYAAKTRTFGNDFOGKYGGAFLDELKAKYPAIFERVQISNGRKLTTN 1118
QY 1043 EKITAKKAYFNGTILGRGVYVLKDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDG 1102
DB 1119 EKITQMSAKYFNGSIQGTGARYIQDNATNOYFVKAGQTFELPKQMTITGS-GPRRVG 1177
QY 1103 NGMTFYSTSGYQAKNSFVQDAKGNWYFDNNGHMVYGLQQLNGEVOYFSLNSGVQLRESFL 1162
DB 1178 DDVQYLSIGGYLAKNTFIQVGANQWYFDKNGNMVTEGEQVIDGKKYFFLDNGLQLRHVLR 1237
QY 1163 ENADGSKNYFHLGNRYNSGYSPNDSSK-WRYFDASGVMAVGLKTINGNTQYFDO-DGY 1220
DB 1238 QGSDGHVYVDPKGVQAFNGFYDFAGPRQDVRYPFGNGQMYRGLHDMYGTTFYDFEKTGI 1297
QY 1221 QVKGAWITGSDCKKBYFDDGSGNMAVNRFA-NDKNGDWYLLNSDGLALVGQTINGKTY 1279
DB 1298 QAKDFIIRPADGRTYFIPDTGNLAVNFAQNPENKAWYLLDSNGYAVTGLQTINGKQY 1357
QY 1280 FQDQKQIKGKIITONGKLYEL-ANSGLARNIFATDSQ----- 1318
DB 1358 FQNEGRQVKGHFVTINQ-RYFLDGDGSEIAPSRFTENNKNWYVDGNGKLVKGAQVING 1416
QY 1319 -----NNWYFSGDGVAVTGSQTIAG 1339
DB 1417 NHYFNNDSYQVKGAWANGRYVDGSGQAVSNQFIQIAANQAWYLNQDGHKVTGLQINN 1476
QY 1340 KKLYPASDCKOVKGSFVTYNGKVHYHADSGLQVNRFEADKGNWYLLDSNGEALTGSQ 1399
DB 1477 KYTYFGSNGAQKGLLTVQGGKKYFDAQHTGQVNRVFAAARGCWYFNSAGQAVTGOQ 1536
QY 1400 RINDQRFVFTREGKQVKGDVAY 1421
DB 1537 VINGKLYFDGSGRQVKGRYVY 1558

RESULT 7

GTFC_STRMU STANDARD; PRT: 1455 AA.
ID GTFC STRMU STANDARD; PRT: 1455 AA.
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-SI precursor (BC 2.4.1.5) (GTF-SI)
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).
GN Name=gTfC; OrderedlocusNames=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gTfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]

Qy	260	PKDILKNGKWTWASSEDRLPILMSWPKQTOIAYLYNMQOGLGTGENYTADSQBSL	319
Db	274	PKYILKDGKWTQSTQEKDFRLLTWTWPDQETQRYVYVYMAQA-LGIHTYNTATSPQL	332
Qy	320	NLAAGTQVQKIEKISOTQOTQWLADIINSFKVTQPNWNSQTESDSAGEKDHLOGGALL	379
Db	333	NLAAGTQVQKIEKISOTQOTQWLADIINSFKVTQPNWNSQTESDSAGEKDHLOGGALL	388
Qy	380	YNSDK-TAYANSYRLLNRTPTSGK--PKYFEDNSSGGYDFLLANDIDNSNPVVOAE	436
Db	389	YNSDK-TAYANSYRLLNRTPTSGK--PKYFEDNSSGGYDFLLANDIDNSNPVVOAE	448
Qy	437	QLNLHLYLNYGSIIVANDPEANFDGVRVDVADVNADLLOIASDYLKAHYGVKDKSEKNAI	496
Db	449	QLNLHLYLNYGSIIVANDPEANFDGVRVDVADVNADLLOIASDYLKAHYGVKDKSEKNAI	508
Qy	497	NHLSLEANSNDPQVNDKTKGAQLPIDNKLRLSLIYALTRPLEKXASNKIERSGLEPV	556
Db	509	NHLSLEANSNDPQVNDKTKGAQLPIDNKLRLSLIYALTRPLEKXASNKIERSGLEPV	560
Qy	557	ITNSLNRNRAEGKNSERMAYPIRAHDSVOTVIAKIIKAOLNPKTDGLTFLDELKQA	616
Db	561	ITNSLNRNRAEGKNSERMAYPIRAHDSVOTVIAKIIKAOLNPKTDGLTFLDELKQA	620
Qy	617	FKIYNEDMRQAKKYYTQSNIPYALMLNKSDSITRLYYGDMYSDGQYMATKSPYYDAI	676
Db	621	FKIYNEDMRQAKKYYTQSNIPYALMLNKSDSITRLYYGDMYSDGQYMATKSPYYDAI	680
Qy	677	DYLLKARIKYAAGGQDMKTYVEGDKSHMDWYTGVLTSVGTGANEATDQSEATKQ	736
Db	681	DYLLKARIKYAAGGQDMKTYVEGDKSHMDWYTGVLTSVGTGANEATDQSEATKQ	731
Qy	737	GMAVITSNPNSKLQNDKVIIVNMGAHKNQRYRPLLLTKDGLTSYTSDAAKSLYRKT	796
Db	732	GMAVITSNPNSKLQNDKVIIVNMGAHKNQRYRPLLLTKDGLTSYTSDAAKSLYRKT	791
Qy	797	NDKGELVFDASDTQYLYNPQVSGYLAVVVPVPGASDNDQVYRVAASNKANATQVYESSAL	856
Db	792	NDKGELVFDASDTQYLYNPQVSGYLAVVVPVPGASDNDQVYRVAASNKANATQVYESSAL	851
Qy	857	DSOLIYEGSNFQDFTKQSDYTNKKIAQNVQLFKSWGVTSEMAPQYVSSDGSFLDSI	916
Db	852	DSOLIYEGSNFQDFTKQSDYTNKKIAQNVQLFKSWGVTSEMAPQYVSSDGSFLDSI	911
Qy	917	IQNGYAFEDRYDLAMSKNNKYGQODMINAVKALHKSIGIQTADWVPDQIYNLPGKEVVT	976
Db	912	IQNGYAFEDRYDLAMSKNNKYGQODMINAVKALHKSIGIQTADWVPDQIYNLPGKEVVT	971
Qy	977	ATRVNDYGEYRKDSEIKNTLYAANTKNSGKDYOQKYGGAFLSALAAKPSIFNRTQISNG	1036
Db	972	ATRVNDYGEYRKDSEIKNTLYAANTKNSGKDYOQKYGGAFLSALAAKPSIFNRTQISNG	1031
Qy	1037	KKIDPEKIKTAKKAYFNGTNIILGRGVGVYLDKNASDKYFELKGNQTYLPKQMTN----	1091
Db	1032	KKIDPEKIKTAKKAYFNGTNIILGRGVGVYLDKNASDKYFELKGNQTYLPKQMTN----	1091
Qy	1092	KEASTGVNDGNGMTFVSTGYQAKNSFVODAKGNWYFDNNGHMYVGLQQLNGEYOYFL	1151
Db	1092	KEASTGVNDGNGMTFVSTGYQAKNSFVODAKGNWYFDNNGHMYVGLQQLNGEYOYFL	1150
Qy	1152	SNGVQURESFLENADGSKNYFHLGNRYSNNGYYSFNDNSKWRYPFASGVMAVGLKTINGN	1211
Db	1151	SNGVQURESFLENADGSKNYFHLGNRYSNNGYYSFNDNSKWRYPFASGVMAVGLKTINGN	1207
Qy	1212	TQYFDQDGQVQKAWITGSDGKRYFDGSGNMAVNRPFANDKMGDYHLYNSDGIALVGQV	1271
Db	1208	TQYFDQDGQVQKAWITGSDGKRYFDGSGNMAVNRPFANDKMGDYHLYNSDGIALVGQV	1267
Qy	1272	TINGKTYVFGQDGKQIKGKIITD-NGKLYFLANSBELARNIFATDSQNNWYFSGDQVA	1330
Db	1268	TINGKTYVFGQDGKQIKGKIITD-NGKLYFLANSBELARNIFATDSQNNWYFSGDQVA	1327
Qy	1331	VTGSOTIAGKLYFASDGKQVKGSPVT--YNGKVHYHADSGELQVNRFEADKGNWYILD	1389
Db	1328	VTGSOTIAGKLYFASDGKQVKGSPVT--YNGKVHYHADSGELQVNRFEADKGNWYILD	1387
Qy	1390	SNGEALTGQSQRINDORVFFTRREGKQVKGVDVAYDER	1424
Db	1388	SNGEALTGQSQRINDORVFFTRREGKQVKGVDVAYDER	1420
RESULT 8			
ID	O68542	STRSL PRELIMINARY;	PRT; 1449 AA.
AC	O68542;		
DT	01-AUG-1998	(TremBLrel. 07, Created)	
DT	01-AUG-1998	(TremBLrel. 07, Last sequence update)	
DT	01-JUN-2003	(TremBLrel. 24, Last annotation update)	
DE	Glucosyltransferase N (fragment).		
GN	Name=glfN;		
OS	Streptococcus salivarius.		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=1304;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=V1477;		
RA	Jaffe R.I.;		
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF049609; AAC05156.1; -; Genomic_DNA.		
DR	PIR; T30552; T30552.		
DR	GO; GO:0015740; P:transferase activity; IEA.		
DR	GO; GO:0009250; P:glucan biosynthesis; IEA.		
DR	InterPro; IPR002479; CW binding.		
DR	InterPro; IPR003318; Glyco_hydro_70.		
DR	Pfam; PF01473; CW binding 1; 2.		
DR	Pfam; PF02324; Glyco_hydro_70; 1.		
KW	Transferase.		
FT	NON_TER	1449	1449
SQ	SEQUENCE	1449 AA; 159895 MW; 0700P6D748471BFB	CRC64;
Query Match			
Best Local Similarity 49.8%; Score 3713.5; DB 2; Length 1449;			
Matches 754; Conservative 207; Mismatches 409; Indels 97; Gaps 22;			
Qy	1	METKRYKQHKYKHWTVAV-----ASGLTILGT-----TTLGSSVSAET	41
Db	1	MDBKVHYKMKVKQWVTIAVTGLSLGASVAVSLGNTDGVVQADEHTDATVAIPDITVD	60
Qy	42	EQOTSQVVTQKSEDDKAESSTQD--APKTQKQTEQOQAGSQANVADTSTI----	94
Db	61	GTVSDNTTAAQDPTTAVAAATNDVADQATPTATFDTTDTTVAANAVDTVAATGTDRA	120
Qy	95	--TKETPSQNIIT--QANSDDKIVNTKSEEAQTSERTKQSEEAQTSASQALTOAKAE	150
Db	121	ATTNDTATNDTAVDTNNNTTDTTNNRAATERRATGARRGPT--GGRATPVNGN	178
Qy	151	LTKQRTAAQENKPNVDLAAIPNV-----KQIDGKYIYSGDQPKQKPFALTVNNKV	202
Db	179	TNNANNVTVWNN--DLPATNNVTDGPSHKITNGKQYVVEDDGTIRKNVYLERIGGS	235
Qy	203	LYFDKNTGALTDSQYQF--KQGLTKLNDYTPHNOIVN-----FENTS---LETIDNYVT	253
Db	236	QYFNAETGELSNOKEYRFDKNGGTGSSADSTNTNTVTNGDKNAFYTTDKDIELVDGYFT	295
Qy	254	ADSWYRPKDIKNGKWTWASSEDRLPILMSWPKQTOIAYLYNMQOGLGTGENYTAD	313
Db	296	ANTWYEPKILLKDGKWTASTENDKEPLLTVWPFSAIQASYLYNMKEQGLGTNQTVTSF	355
Qy	314	SGOESLNLAAQTQVQKIEKISOTQOTQWLADIINSFKVTQPNWNSQTESDSAGEKDH	373
Db	356	SSQOTMDQAALEVQKRIEGRAREGTDWLRTIKNFVKTPQGNWSTSE---NLDNNDHL	412
Qy	374	QSGALLYNSDKTAVANSYRLLNRTPTSGK--PKYFEDNSSGGYDFLLANDIDNSNP	431

Db 413 QGALLYNDSTRSHANSYRLNRTPTSTQTKRNPYTKDTNGGFEFLANDINDNSP 472
QY 432 VVQAEQLNLWLHYLMNYGSIIVANDPEANFDGVRVDVAVNNVADLLQIASDYLYKAHYGVDKS 491
Db 473 AVQAEQLNLWLHYIMNIGTITGGSEDFDGVVDVAVNNVADLLQIASDYLYKAHYGVDKS 532
QY 492 EKVAINHLSILEAWSNDPOYNKDTKGAQLPIDNKRLSLLYALTRPLEKDAASKNEIRS 551
Db 533 QDAIHLKLSILEAWSHNDAYNEDTKGAQLPMDPMLHALVYSLRPIGN-----RS 584
QY 552 GLEPVTNSLNNSRAEGKNSRMANYIFIRAHSEVQTVIAKIIKQINPKTDLGLFTFLD 611
Db 585 GVEPLNSLNSDRSESGKSKRMANYAFVRAHSEVQSIIGQIIKNEINPSTGNTFTFLD 644
QY 612 ELKQAFKYNEDMRQAKKYTQSNIPYALMLSNKDSITRLYGYDMYSDGQYMATKSP 671
Db 645 EMKKAPEIYNKDMRSANKQYQYNIPSAVALMTHKDTVPRVYGYDMYTDGQYMAKSP 704
QY 672 YYDAIDTLKARIKYAAGGQDMKITVYEGDKSHMDWDYTOVLTSVRYGTGANATDQGE 731
Db 705 YYDAIETLLKGRIRYAAAGGQDMKVIY-GYGTNGWDAGVLTSVRYGTGANSASDTGTA 763
QY 732 ATYQGMVITTSNNPSIKLNQNDKVIIVNMGAAHKNQOEYREPLLTTKDGLTSYTSDAAKS 791
Db 764 ETRNQGMVITSVNQPALRLTSN--LTINMGAAHKNQOEYREPLLTTKDGLTSYTSDAAKS 821
QY 792 LYRKTNDKGBELVFDASDIQGLYNPQVSGVLAIVPVPVPGASNDQVRVAASKANATGOVYE 851
Db 822 IVKYTDGNGNLTFSEANEIRGNPQVDGYLAIVPVPVPGASENDQVRVAPSKENSSGLVYE 881
QY 852 SSSALDSQLIYEGFSNFPQVTKDSYTNKIIAQNVLFPKSGWVTSFEMAPQVSVSDGS 911
Db 882 SNAALDSQVIYEGFSNFPQVTPNPSQVYTNKIIAENANLFPKSGWITSFEPAPQVSVSDGS 941
QY 912 FLDSIILQNGVAPEDRYDLAMSKNKYGSQODMINAVKALHKSQIYQIADWVPDQIYNLPG 971
Db 942 FLDSVQLNGVAFYDRYDIGNSKNDKYGSLADLKAALKSLHVGISAIADWVPDQIYNLPG 1001
QY 972 KEYVTATRVNDYGEYRKDSIKNTLYAANTKSGNDYQAKYGGAFSLSELAAYPSIFNRT 1031
Db 1002 DEVVTATRVNNGYGETKDGALIDHSLYAATRFPGNDYQKYGAFDELKRLYQIYFDRV 1061
QY 1032 QISNGKKIDPSEKIIATWAKYFNGTNIILGRGVYVILKNASDKYFELKGNQTYLPRQMTN 1091
Db 1062 QISTGKRMTTDEKIIKWSAKYMNGTNILDGSEYVLK-NGLNGYGTNGKVSLSLPRVGS 1120
QY 1092 KEASTGFVNDGNG-----MTFYSTSGYQAKNSFVQDAGKNWYIFDNGHMYGLQ 1141
Db 1121 NOSTNGDNQNGDGSKGFEKELFSVRYRNNQYAKNAFTKDNNGVYIFDNGSRMAVGSK 1180
QY 1142 QLNGEVQYFSLNGVQLRSEGFLENADGSKNYFGLHGNRYNGYV--SFDNDSK----- 1191
Db 1181 TIDGKQYFLANGVQLRDGYRQNRQVYFYDQNGVLSANGQDPKPDNNNTSGRNPV 1240
QY 1192 -----WRYFDASGVMAVLKTINGNTQYFDQNGYQVKGAWITGSDGKKRYFDDGSGNMA 1245
Db 1241 QIGNNWYAYDNGKRVGHONINGQELFFDNGGVQVKGRTV-NENGALRYIDANSGEWA 1299
QY 1246 VNRFPANDKNGDWYILNSDGIALVGVTINGKTYIFGQDGKQIKGIITDNGKLYPLANS 1305
Db 1300 RNRFPABIEPFWAYFNNNDTAVKGSQNINGQDLVYFDQNGRQVKGALANVYDGNLRYDVS 1359
QY 1306 GELARNIPATDSQNNWYIFGSDGVAVTGSQTAGKLYFASDGKQVKSFTVYNGKVVHY 1365
Db 1360 GELYNRFP-HEIDGSGWYIFDNGNAVKGMWNINGQMLLFDNGKQIKGHLVRVNGVRYF 1418
QY 1366 HADSGELVNRFPADKNGWYILDSNG 1392
Db 1419 DPNSGEMAVNRVVEVSPGWVYFDGEG 1445

ID Q55264_STRSL PRELIMINARY; PRT; 1449 AA.
AC Q55264;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Name=gtfL;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL; LJ5495; AAC1412.1; -; Genomic_DNA.
DR FIR; T30857; T30857.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
FT SIGNAL 1 35 Potential
FT CHAIN 36 1449 Glucosyltransferase.
SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 49.8%; Score 3713.5; DB 2; Length 1449;
Best Local Similarity 51.4%; Pred. No. 4.4e-148;
Matches 754; Conservative 207; Mismatches 409; Indels 97; Gaps 22;

QY 1 METKRYKQHKHKKHVVAV-----ASGLITLTG-----TTLGSSVSAET 41
Db 1 MDKRVHYKHKVKKQWTTIAVGLSLGAVSAVSLGTNDGVQADEHTDATVAIPDITVD 60
QY 42 EQGTSKVVYTKSEDDKAESSQTD--APKTKQAQTEQTAQSQANVADTSTSI----- 94
Db 61 GTVSNDDTTAAQDPTTAAVAAATNDVATDQATPTATFDLTDTTNTVAANAVDTVATGTDRA 120
QY 95 --TKETPSQNIIT--QANSDDKTVNTKSEEAQTSERTKQSEEAQTASSQAALTQAKAE 150
Db 121 ATNTDQATNDTAVDTTNNNTTDTTVDRAATERRATGARGPT--GGRATPVGN 178
QY 151 LTKQRTAAQENKNPVDLAAIPNV-----KQIDGKYYIGSDGQPKKFNALTNNKV 202
Db 179 TNNANNTVTVNN--DLPATNNVTDGSPSHIKTKINGKQYVVEDDGTIRKNVYLERIGGS 235
QY 203 LYFDKNTGALDTSQYOP-KQGLTKLNNDYTPHNQIVN-----FENTS---LETINYVT 253
Db 236 QYFNAETGELSNOKEYRFDKNGTGSSADSTNTNTVNGDKNAFYGTDDKIDELVDGYFT 295
QY 254 ADSWYRPKDIKNGKTTWATASSEDLRPLILMSWPDKQTOIAYLYNNVQOGLGTGENYTD 313
Db 296 ANTWTRPEILKDGKEWTASTENDKRPLLTVMWPSKAIQASLYNMYKEOGLGTNYTYSF 355
QY 314 SSQESLNLAQAQTVQVKIETKISQTOQTOWLRDIINSFVKTPQNNWSQTSSTDSAGEKHL 373
Db 356 SSQTDQDAALAEVQKRIEERAREGNTDLRTTIKNFVKTPQCNWSTSE--NLDNDHL 412
QY 374 QGALLYNSDKTAYANSYRLNLTPTSTQTK--PKYFEDNSSGGYDFLLANDINDNSP 431
Db 413 QGALLYNNDSRTSHANSYRLNLTPTSTQTKRNPYTKDTNSGGEFFLLANDINDNSP 472
QY 432 VVQAEQLNLWLHYLMNYGSIIVANDPEANFDGVRVDVAVNNVADLLQIASDYLYKAHYGVDKS 491
Db 473 AVQAEQLNLWLHYIMNIGTITGGSEDFDGVVDVAVNNVADLLQIASDYLYKAHYGVDKS 532
QY 492 EKVAINHLSILEAWSNDPOYNKDTKGAQLPIDNKRLSLLYALTRPLEKDAASKNEIRS 551

Db 533 QDAQIKHLSILEAWSHNDAYNEDTKGAQLPMDPMHLALVYSLLRPIGN-----RS 584
Qy 552 GLEPVTJNSLNRASBKNSERMANYIFIRAHSEVOTVIKIIKAIQNPKTDLFTFTLD 611
Db 585 GVEPLISLNSLDRSESGKSKRMANYAFVRAHSEVQSIIGQIIKNEINPQSTGNTFTLD 644
Qy 612 ELKQAFKINEDRQAKKYTQSNIPITAYALMNSKDSITRLYGYGMDSDGQYMATKSP 671
Db 645 EMKQAFKINEDRQAKKYTQSNIPITAYALMNSKDSITRLYGYGMDSDGQYMATKSP 704
Qy 672 YYDAITLKLARIKYAAGGDMKITVVEGDKSHMDWDTGVLTSVRYGTGANEAATQDQSE 731
Db 705 YYDAITLKLARIKYAAGGDMKITVVEGDKSHMDWDTGVLTSVRYGTGANEAATQDQSE 763
Qy 732 ATKTQGMVITSNPNSLKLNDQKVIUNGAHKNQYRPLLTTKDGLTSYTSDAAKS 791
Db 764 ETRNQGMVITSNPNSLKLNDQKVIUNGAHKNQYRPLLTTKDGLTSYTSDAAKS 821
Qy 792 LYRKTNDKGLVFDASDIQGYLNPQVSGYLAVMPVPGASDNODVRVAASNKANATQVYE 851
Db 822 IVKYTDGNGNLTFSAHEIRGNPQVDGYLAVMPVPGASENQDVRVAPSKERNSSGLVYE 881
Qy 852 SSSALDSQLIYEGFSNFQFVTKDSYTNKKIAQNVOLFKSMGVTFSFEMAPQVSSDDGS 911
Db 882 SNAALDSQVYEGFSNFQFVTKDSYTNKKIAQNVOLFKSMGVTFSFEMAPQVSSDDGS 941
Qy 912 FLDSIIQNGYAFEDRYDLAWSKNNKYGSQODMINAVKALHKSQIOVIADWVDPQIYNLPG 971
Db 942 FLDSVIQNGYAFEDRYDLAWSKNNKYGSQODMINAVKALHKSQIOVIADWVDPQIYNLPG 1001
Qy 972 KEVWTATRVNDYGEYKDSIKNTLYAANTKSNKDYQAKYGGAFSELAAYPSIFNKT 1031
Db 1002 DEVVTATRVNNGYETKDAIIDSLSYAAKTRTFGNDYQGYGAFDELKRLYPOIFDRV 1061
Qy 1032 QISNGKKIDSEKITAUKAKYFNGTILGRGVYVLKDNASDKYFELKGNQTYLPRKMTN 1091
Db 1062 QISTGKRWITDEKLTQWAKYNGTILGRGVYVLKDNASDKYFELKGNQTYLPRKMTN 1120
Qy 1092 KEASTGFVNDGNG-----MTFYSTSGYQAKNSFVQDAKGNVYFDDNNGHMYGLQ 1141
Db 1121 NQSTNGDNQNGDGSKFELFSVRYRNNQYAKNAFIKNDNGVYFDDNNGHMYGLQ 1180
Qy 1142 QINGEVQYFLSNGVOLRESFLENADGSKYFHLGNRYNGY--SFNDNSK----- 1191
Db 1181 TIDGKQYFFLANGVQLRDGYRQNRGQVYFDQNGVLNANGKQDPKPDNNNNSGRNQFV 1240
Qy 1192 -----WRYFDASGVMAVLKTINGNTQVFDQNGVQKGAWITGSDGKKRYFDDGSGNMA 1245
Db 1241 QIGNWYAYDNGKRVTHQNINGQELFFDNGVQVKGRTV--NENGAIYYDANGGEWA 1299
Qy 1246 VNRFANDKNGDYYLNSDGIALVGVQTINGKTYFFQDQKQIKGKIITDNGKLYFLANS 1305
Db 1300 RNRFAETEPGWAYFNNDGTAIVKGSQINGQDLYFDQNGRQVKGALANVDGNLRYDVNS 1359
Qy 1306 GELARINPATDSQNNYFSGDGVAVTGSQTIAGKLYFASDCGVKGSFVTVNGKVHY 1365
Db 1360 GELYRNR--HEIDGWSYTFDNGNAVKGWYNGNQNLFFDNGKQIKGHLVRVNGVRYF 1418
Qy 1366 HADSGELQVNRFEADKQGNWYLLDSNG 1392
Db 1419 DPNSGEMAVNRVVEFGWVYFDGEG 1445

RESULT 10
Q9WXJ5_9STRE
ID Q9WXJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5_9STRE
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE GTF-S.
GN Names=gtf;
OS Streptococcus criceti.

Qy 1 METKRYKMHKVKCHWTVAVAS-GLI-----TLGTTTLGSSVSAETEQOTS--- 46
Db 1 MERKRYKMHKVKCHWTVAVAS-GLI-----TLGTTTLGSSVSAETEQOTS--- 60
Qy 47 DKVVTQ-----KSEDDKAASESSQTDAPKTKQAQTEQTAQSQANVADTSTSIKTPSQN 102
Db 61 NKEVDSDAQAATDAKATSEQAASVSDT--NSETDQVQ-----NVDGVSANNQPEAPA-- 112
Qy 103 ITTOANSDDKTVNTYKSEBAQTSSEBTQKSEBAQTSSEBAQTSSEBAQTSSEBAQTSSEBAQTS 162
Db 113 -APQAAASNTTAT--SEANTN---TAVSEAA--PAAENRTAEBKADLSQDEEAAA--- 162
Qy 163 KNPVDLAAIPNKVDIGKYYIIGSDGPKKNEFALTNNKLVYFDKNTGALTDTTSQYQFQK 222
Db 163 -----LSLNDIKKVDGKYYIIGSDGPKKNEFALTNNKLVYFDKNTGALTDTTSQYQFQK 216
Qy 223 GLTKLNNDYTPHNQIVNFPNTSLETIDNTYVADSWYRPRKDIKNGKWTWASSSDLRPL 282
Db 217 GLTPLVSDPS INKAFDSSAKSFELVGYLTAAESWYRPTKILKNGKWTWASSSDLRPL 276
Qy 283 MSWVDPKQTOIAYLNNYQOGLTGENTYADSSQSLNLAQTVQVVKIETKISOTQOTOW 342
Db 277 TSWVDPKQTOIAYLNNYQOGLTGENTYADSSQSLNLAQTVQVVKIETKISOTQOTOW 335
Qy 343 LRDIINSFVKTPNNNSQTESDTSAGEKDHLOGGALLYNSDKTAYANSDYRLNRTPTS 402
Db 336 LRDAWAAFAVATOSRWNSDSE---QPKNDHLOGGALLYNNKLTWADSKYRLNRTPTS 392
Qy 403 QTKPKPYFEDNSGGYDFLLANDIDNSNPVQAQQLNHLHYLMYNSGSIIVANDPEANPDGV 462
Db 393 QDGKTHYKADVEYGGYFLLANDVDSNPVQAQQLNHLHYLMYNSGSIIVANDPEANPDGV 452
Qy 463 RVDADVNNADLIQATSDYLAHYGVYKSEKAINHLSILEAWSNDNDPOYKDKTKQAOLP 522
Db 453 RVDADVNNADLIQATSDYLAHYGVYKSEKAINHLSILEAWSNDNDPOYKDKTKQAOLP 512
Qy 523 IDNKLRLSLYALTRPLEKXDNKNEIRSGLEPVTN--SLNRSAGKNSERMANYIFI 580
Db 513 MUNGRLSLYALTRPLEKXDNKNEIRSGLEPVTN--SLNRSAGKNSERMANYIFI 565
Qy 581 RAHDSVQTVIAKIIKAIQNPKTDLFTLDELKQAFKIYNEDMRQAKKYTQSNIPTAY 640
Db 566 RAHDSVQTVIAKIIKAIQNPKTDLFTLDELKQAFKIYNEDMRQAKKYTQSNIPTAY 625
Qy 641 ALMLSNKDSITRLYGYGMDSDGQYMATKSPYDAITDGLLKLARIKYAAGGDMK----- 694
Db 626 AVMLSNMESVTRVYVYGDFTDQGYMETKSPYDAITDGLLKLARIKYAAGGDMK----- 685
Qy 695 --ITYVEGDKSHMDWDTGVLTSVRYGTGANEAATQ--GSEATKTQGMVITSNPNSLKL 751

OG Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSSP; P06654; 1MPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 49.1%; Score 3661; DB 2; Length 1512;
Best Local Similarity 49.3%; Pred. No. 7.6e-146;
Matches 759; Conservative 230; Mismatches 400; Indels 150; Gaps 31;

Db 686 SAAMKAKNPDSGSLGNSEVLVSRFGQDVMSADMTGGQAKTSGMFSLIANNPELELD 745
Qy 752 QNDKVIIVMGAHAKNQYRPLLLTTKQGLTSYTDAAAKSLYRKTNDKGLVFPDASDIQ 811
Db 746 ANBEIKVNVGKIHAQAYRPLLLTTKGLQKYLNDSDT-NUTKVADKQDFTTFKGSIKG 804
Qy 812 YLNFQVSGYLAVWVPVGSNDQDVRVAASNKANATG-QVYESSALDSQLIYEGFSNPD 870
Db 805 YKQVEVNGYLSVWVPVGSNDQDVRVAASNKANATG-QVYESSALDSQLIYEGFSNPD 864
Qy 871 FVTKDSYTNKKIAQNVOLFKSWGTSFEMAPQVVSSEDSFLLSIIONGVAFEDRDYLA 930
Db 865 FVQKDAQYTNKKIAENTDLFKWGVTSFEMAPQVVSATDGFLLSIIONGVAFEDRDYLA 924
Qy 931 MSKNKYGSGQDQMINAVKALHSGIQVIADWVPDQIYNLPGKEVVTATRVNDYGEYRKDS 990
Db 925 MSKNKYGSKEDLANALKALHAAGIQALADWVPDQIYQLPGKEVVTASRVNDYGRVKIDQ 984
Qy 991 EIKNTLYAANTKSGKDYQAKYGAFLSELAAKYPSIPNRTQISNGKKIDPSEKITAWKA 1050
Db 985 PMVKNLYLANTKSSGKDFQAKYGGFLAELQKQYPEMFTAKMISTGKPIDSSVKLEWSA 1044
Qy 1051 KYFNGTILARGVGYVLKDNASDKYFELKGNQTYLPKQMT-NKEASTGTFVNDGNGMTFYS 1109
Db 1045 QYFNGTIVLGRGTDYVLSDEGTGKYFTVNEKGEFLPAVLTDGKEAKTGFYNDGKMTYFT 1104
Qy 1110 TSGYQAKNSFVQAKGNWYFDNNHMYVGLQQLNGEVBQVFLSNGVOLRESFLENADGSK 1169
Db 1105 TAGSOAKSDFVTVAGNTYFDYTGHWVTGPGINTKFFYFLPNCVWLKDAVWEDDRGRS 1163
Qy 1170 NYFGLHGRY-----SNGYISFDNDSK-----WRYFDASGVMAVGLKTINGNTQYFDQDGYQ 1221
Db 1164 VYTGKTVMYKGSNNWFAM-TDSKGQLRFRHFNDYGFMSVGLVTHGNVQIYDEBGFQ 1222
Qy 1222 VKGAWITSGDKRYKYPDGGSNVAVRPFANDKNGDWYILNSDGLALVGVOTINGKTYFP- 1280
Db 1223 VKGDFVTDKAGQTRYFDKNTGNLVKGQFFN-QNGHWYISDDQGLIAKGAQTIKQKLYFD 1281
Qy 1281 GQDQKQIKGKIITD-----NG 1296
Db 1282 AKTGAQVKGDFVTDKQNTFFYSGDTCGLAVSTFFSTGNNAWFYADENGHVAKEKTING 1341
Qy 1297 KLKYFLANGSELARNIPATDSQ-----NNWYIFSGDGVAVT 1332
Db 1342 QKLYFDTKTQQAQGRFVRDAKGLRFDYADTGALVTNSFLETRAGSNQWYMGADGVAVR 1401
Qy 1333 GSOTIAGKLYFP-ASDGKQVKGSEFT-YNGKVHYHYHDSGELQVNRFEADKQGNWYLD 1390
Db 1402 GHQTIOSRHHYFPAETGQQAQGIWVTDANGRKTYFDANTGDRVNVQFVL-VNGSWYFFGY 1460
Qy 1391 NGELATGSORINQORVFFFTREGKQVGDVAYDERRLLVY 1429
Db 1461 DGAATVGFDRIRQHLVFNPDGTQAKGTTVKIDNRIYTF 1499

RESULT 11

Q56CX8_9STRE PRELIMINARY; PRT; 1506 AA.
AC Q56CX8;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DT Glucosyltransferase-T.
GN Name-gtT;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY966490; AAX76986.1; -; Genomic_DNA.
KW Transference.
SQ SEQUENCE 1506 AA; 167016 MW; 948882EF2C17C451 CRC64;
Query Match 48.8%; Score 3639; DB 2; Length 1506;
Best Local Similarity 49.3%; Pred. No. 6.4e-145;
Matches 724; Conservative 247; Mismatches 403; Indels 96; Gaps 26;
Qy 1 METKRRYKMKHKKHWVTAVAS-GLITL-GTTTLGSSVSA-----ETEQOTSQVVTQK 53
Db 1 MERKLYKMKHKKQWVTIAVASAGLASIVGAGLSQTVSADDLAKQAASQQAANAQ 60
Qy 54 SEDDKAASESSQTDAPKTKQ-----AQTEQTAQSQANVADTSTSTIKETPSQNI 103
Db 61 NEDEVASDAADTASAKATSEKVVQSSDNTSETNOVETKQASAKESADAVAKAP----- 116
Qy 104 TQANSDDKTVNTYKSEEAQTSERTKQSEEAQTTASSQALTOAKABLTQKQRTAAQENK 163
Db 117 -----QAGPATTSQ--VASSSESSVAPSKADKAAAGSVSQNEEAAA----- 156
Qy 164 NPVDLAAIPNVKOIDGKYIYIGSDGQPKKQNFALTNNKVLVYFDKNTGALTDTTSQYQKQ 223
Db 157 -----LSLANIKKIDGKIYVMADGSKYKQNFALTVDGQMLYFDAKTGALSTSTYSQ 211
Qy 224 LTKLNNDYTPHNOIYNFENTSLTIDNYVTADSWYRPKIDILKNGKWTASSESLRPLM 283
Db 212 LTPVSDPSVNNKAFDSSEKSFELVDGYLTAESWYRPAKILENGKTWVDSKETDLRPVL 271
Qy 284 SWPDKQTOIAYLNNYNNQGLGTGENYTDSSGESLAAQTVQVKIETKISQTOQOWL 343
Db 272 SWPDKQTOVAYLNNYS-KALGKKEFTTETSQTLNTAAELIQAIEARVSKEQTKWL 330
Qy 344 RDIINSFVKTOPNWSQTESDTSAGEKDHLOGGALLYNSDKTAYANSYRLLNRTTSQ 403
Db 331 REAATAFVQSRWNKDSQYDKA---DHLQGGALLYTNNNLTSEWANSRLLNRTTRQ 387
Qy 404 TGPKPTFEDNSSGGYDFLLANDIDNSNPVVOAQLNWLHYLMVYGSIVANDPEANPGVR 463
Db 388 DGKTHYSKADKYGGYEFLLANDVNSNPVVOAQLNWLHYLMVYGSIVANDPEANPGVR 447
Qy 464 VDAVDNNDLQIASDYLKAHYVDKSEKNAIHLISILEASDNDPQYNKDTKGAQLPI 523
Db 448 VDAVDNNDLQIATYNYFNSVYGVNKAQAALAHISVLEAWSYNDNDYNDQTINGAALAM 507
Qy 524 DNKLRLSLAYALTRPLEKASNKNEIRSGLEPVTN--SLNPRSAGKNSERMANIYFIR 581
Db 508 DNGLRFSLLYTLRPT-----NERTPGMSTLIKSEYGLTDRTKNDKYGDTQPSYFVR 560
Qy 582 AHDSEVQTVIAKIIKAQINPKVTDGLTFLDELKQAFKIYNEDMRQAKCKYTQSNIPATA 641
Db 561 AHDSEVQTVIAKIIKEKIDPTDGTFTLDLQKQAFELYNKDNVSNKHYTHYINIPATA 620
Qy 642 LMSLNKDSITRLYYGDMYSDGQYMAKSPYDAIDTLKARIKYAAGQDMKITVYEGD 701
Db 621 VMLSNNMESVTRVYVYGLTDDGQYMAKSPYDAINTLLRARIRYAAGQIMEHNSYKPS 680
Qy 702 -----KSHMD-----WDYTGVLTSVRYTGANEATDQ--GSEATKTQGMNAVITSNP 752
Db 681 AAMKAHPDAGNVLGNSEVLVSRFGQDVMSADMTGGKLAKTSGMFTLISNNPELELDV 740
Qy 753 NDKVIIVMGAHAKNQYRPLLLTTKQGLTSYTDAAAKSLYRKTNDKGLVFPDASDIQ 812
Db 741 NEEIKVNVGKIHAQAYRPLLLTTKGLQKYLNDSDT-LTKIADKQDFTTFKGSSEIKGY 799
Qy 813 LNPQVSGYLAVWVPVGSNDQDVRVAASNKANA-TQVYESSALDSQLIYEGFSNFQDF 871
Db 800 KQVEVNGYLSVWVPVGSNDQDVRVAASNKANA-TQVYESSALDSQLIYEGFSNFQDF 859
Qy 872 VTKDSYTNKKIAQNVOLFKSWGTSFEMAPQVVSSEDSFLLSIIONGVAFEDRDYLA 931
Db 860 VQKDSYTNKKIAENTDLFKWGVTSFEMAPQVVSATDGFLLSIIONGVAFEDRDYLA 919
Qy 932 SKNNKYGSGQDQMINAVKALHSGIQVIADWVPDQIYNLPGKEVVTATRVNDYGEYRKDS 991

	: : : : : :	
920	SXNKYGSKEDLANALKALHAAGIQAIAADWVPDQIQOLPGKEVVTSARVNYGRVKVDQP	979
Qy	I KNTLYAANTKSNGKDYOAKYGGAFLSELAAKPSIENRTQISNGKKIDPSEKITAWKAK	1051
Dd	: : : : : : :	
980	LVEKLVLANTKSSGKDFOAKYGGEBFLAELOKKYPPEMTTQIWISTGKTIDISVKLKEWSAK	1039
Qy	YFNGTNILRGVGYYLVLDKNASDKYFELKGNTYLPKOMT-NKEASTGFVNDGNMTPYST	1110
Dd	: : : : : : :	
1040	YFNGTNVLDRGTDYILSDEGTGYFTVNEKGFPLPASLTGNKDAKTFYNDGKGIVYTT	1099
Qy	SGVOAKNSFVQDAGKNWYFDNNGHMYGIGLQOLNGEVOYFSLNGVOLRESFLENADGSKN	1170
Dd	: : : : : : :	
1100	AGNKARSAPFTEA-GNTYFYDYTHMVTPGNVINTKFYFPLNGIMLKDAIKODEKRSV	1158
Qy	YFGHLGNRYSNGYYSPFNDNSKW-----RYPDASGVMAVGLKTINGNTQYEPDQ	1218
Dd	: : : : : : :	
1159	YYGKTGMVYKGG-----RDNEWPAMTDSKQMRFPHEDRYGPMSIGLVTINQNQVYDEN	1213
Qy	GYYQVKGAWITGSDGKKRYFPDDGSGNMVNRPANDKGDWYYLNSDGIALVGVQTINKTY	1278
Dd	: : : : : : :	
1214	GFQVKGFEVTDQDGOTRYFDOGSGNLVKGOFLN-KDGNWYYLDDOGILAVKAQAOTIKGOQL	1272
Qy	YFG-QDCQOIKGKIITD-NGKLYFLANSSELARNIEATDSQNNWYFEGSDGVAVTGSOT	1336
Dd	: : : : : : :	
1273	YFDTKTGQVKGDFVTDKOGNTPFYSGBTDLIIQGFFSTGNNAWFYADENGHWKAGAKT	1332
Qy	IAGKKLYF-ASDGKVQRGVSFVTVNGKYHYHADSGELQVNRFEADKDN--WYYLDSNGE	1393
Dd	: : : : : : :	
1333	IRGKLYFDTKTGQAAGRFRIDDKGVRYDYADDTGLVTNAFLBETKAGSNQWYYMGADGY	1392
Qy	ALTGQSQRINDORVFETRE-GKOVKGDVAYD	1422
Dd	: : : : : : :	
1393	AVKGNQTIKQNHWFDAETCGQAAGIITVD	1422

RESULT 12

```

08KZL5_95STRE PRELIMINARY;      PRT; 1554 AA.
Q8KZL5_95STRE
AC Q8KZL5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name-gtfu;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1310;
[1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=2195868; PubMed=11960691; DOI=10.1016/S0304-4165(01)00240-9;
RA Hanada N., Fukushima K., Nomura Y., Sempuku H., Hayakawa M.,
RA Mukasa H., Shiroza T., Abiko Y.;
RT "Cloning and nucleotide sequence analysis of the Streptococcus
RT sobrinus gtfu gene that produces a highly branched water-soluble
RT glucan."
RL Blochim. Biophys. Acta 1570:75-79(2002).
DR EMBL; AB089438; BAC07265.1; -; Genomic_DNA.
DR HSP; P06653; 1H8G.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Transferase.
SQ SEQUENCE 1554 AA; 171676 MW; 6981BCC1DAE24A73 CRC64;

```

Qy	1	METKRYKMHVKVGHVWTVAVAS-GLITL---GTTTLGSSVSAETEQTSDKVVTQKSD 56
Db	1	MEKLLHYKLHKVKKHVVTTIAVASIGLVSIVGAGTVAEDKVANDTTAQATVGVDTGQ--- 57
Qy	57	DKAASSSTQDAPKT---KQATEQTOAQs-QANVADTSTSTTKETPSQNIITQANSDD 111
Db	58	DQATTTDANTNTTDTTDAQSANTNOQAGSDQSNNOQAKQDT-----ANTDR 106
Qy	112	KTVNTNKSBEAQSTKQSEBAQTTASSQAULTQAKAELTKORQTAAQKKNKPNVDLAAI 171
Db	107	NQADNSQTDNNQATDQATSPATDGTSVQRDAANVATA-ADQSGQTAPSEQKSAAL-SL 164
Qy	172	PNVKQIDGKYYIIGSDGQPKNFALTNNKVLVPDKNTGALTDTSTQYQFGUHLTKLANDY 231
Db	165	DNVKLIDGKYYVQADGSKKFNALTNGOMLYFSDTGALSSSTYSTYSFSGTTLNVDDF 224
Qy	232	TPHQIWNENTSLFTIDNYVTADSWRPKDILKNGKTTWTTASSEDLRPLLMSSWPDQOT 291
Db	225	SSINKAYDSTAKSFELVNGYLTANSWYRPAGILRNQGTWEASNENDLRPVLMSWPDQOT 284
Qy	292	QIAYLNMNCOGLGTGENTYTDSSQBSLNLAAOTVQVKIETKISQTOQTQWLDRDINSFV 351
Db	285	QVAYVNMKYLANSETEVTNETSQVDLNKEAQSIQTKIEOKLTSDNSQTOWLRATMAEAFV 344
Qy	352	KTOPNNSQTESDTSAGEKDHLQGGALLYSNSDKTAYANSDYRLLNRTPTSTQTKPKPYFE 411
Db	345	AAQPKNMWSTE---NFNKGDLQGGALLYTNSDLTPWANSDYRLLNRTPTQDQTKKYFT 401
Qy	412	DNSGGYDFLLANDINDSNPVPVQAEQLNWLHYLMNYGSIIVANDPEANFDGVRDADVNDVN 471
Db	402	EGGEGYEFLNSDNDVNSNPVPVQAEQLNQHLYLMNWGDIIVMGDKDANFDGVRDADVNDVN 461
Qy	472	ADLLOJASDYLKAHYGVYDSEKNAIINHLSILEAWSNDNDPOYKNDTGAQLPIDNKLRLSL 531
Db	462	ADLQWISNFKDNYKVYUJSEANALAHISILEAWSLNDQNYEDTNGTALSINDSSRLTS 521
Qy	532	LYALTR-PUK-DA-SKNKEIRSGLEPVTNLSL-NR-SEAEGKNSERMANYIFIRAHDSVQ 588
Db	522	LAVLTQPGQRIDLN-----LISESVKERANDTAYGDTIPTYSFVRAHDSVQ 571
Qy	589	TVIAKIIKQINPKTDLFTLDELKQAFKIYNEDMRQAKKKTQSNIPTAYALMLSNKD 648
Db	572	TVIAKIVKEKIDTNSDGYTFLDQLDAFKIYNEDMAKNVKTTHYNIPAPYALLLSNME 631
Qy	649	SITRLYYGDMSYSDGQYMAKTSPYDAIDTLKARIKYAAGQDMKITTYEGDKSHMDWD 708
Db	632	SVPRVYGGDYTDGQYMAKTSPYDAIATMLQRIAYVSGGSEEVHKVNGNQ----- 686
Qy	709	YTCGLTSVRYGTGANEATD-QGSEATKTQGMVITSNPSLKLQNQDKVIVNMGAAHKQ 767
Db	687	---ILSSVRYGQDLMSADDTQGTDLSTGLVTLVSNDPNLDLG-GBSLTVNMGRAHANQ 742
Qy	768	EYRPLLTTHKGLTSTYSDDAAKSLRYKTNWKGELVPDASDIQGYLNPQVSGYLAVWVPV 827
Db	743	AYRPLILGTGQVGYLKDSDT-NIVKYTDANGNLTFADDIUKGYSTVDMSGYLAWVPV 801
Qy	828	GASDNODRVAA-SNKANATQVYESSALDSOLITYEGFSNFDQPVTKDSYVTNKKIAQNV 887
Db	802	GAKDQDQVRAADTNQKADGSKLSTSAALDSQVIYEGFSNFDQFANNADATYTNKKIAENA 861
Qy	888	QLFKSGVTSFEMAPQVSSDEGSLDTSITQNGYAFEDRYDLAMSKNNKTKGSOQDMINAV 947
Db	862	DFPKLGITSFEMAPQVVSATDGSFLDSITQNGYAFSDRYDLAMSKNNKYGSKDDLANAL 921
Qy	948	KALHKGGIQVIADWVPDQIYNLPGKEVVTATRVNDYGEYRKDSIEKTLVAANTYKSGKD 1007
Db	922	KALHANGIQAIADWVPDQIYQLPGEVVTAKRTNSYGNPTFDAYINNALLYATNTKSSGSD 981
Qy	1008	YQAKYGGAFSLSEAAKYPSIFNRQTQSNKKIDPSEKITAKWKAFNGTTLGRGVGVVL 1067
Db	982	YQAKYGGAFULDELKAKYFDMFTVMNLSTGPIPDSTKIKQWEAKYFNGTINVLGKAGTVL 1041
Qy	1068	KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMFTFYSTSGYQAKNSFYQDAGKN 1126


```
Db 1042 SDDATGKYFTVWNGDFLPASFTGDQNAKTFYDGTGMAYYSTSGNKAVNSFIYEG-GH 1100
Qy 1127 WYTFDNNHVMYGLQQL-NGEVOYFLSNGVOLRESFLENADGSKNYFCHLGNRY-SNGY 1184
Db 1101 YYYFDKDHVMTGSYKAEDGNDYYFLPNGLQMRDAIYQDAQNSYYVGRGILYKGDNWY 1160
Qy 1185 SF-----DNDKWRYPDAGMVAVLKTINGNTQYFDQDGYQVKGAWITGSDGKRYFDD 1239
Db 1161 PFVDPNNANKTVFRYPDANNVMAIGYRNMVGTYYFDENGFOAKGQLLTDKG-THYFDE 1219
Qy 1240 GSGMVAWRFANDKNGDWYLLNSDGLALGVQTINGKTYFYG-ODGKQIKGLITD-NGK 1297
Db 1220 DNGAMAKNGKFN-VGDDWYIMDNGNAVKQYPPVNNQILYFNPETGVQVKGQFITDAQGR 1278
Qy 1298 LKTYFLANSBELARNIFAT-----DSQNNWYFG-SDGVAVTGS 1334
Db 1279 TSYDANSALKSGSFTTPNGSDWYEAENGVYVKGFKQVAENQDQYFDTTGGKQAKGA 1338
Qy 1335 OTIAGKLLYPASP-GKQVKSFTV-YNGKVHYHADSGELQVNRFEADKGNWYLLDSNG 1392
Db 1339 AKVDGRDLYPNPDSGVQVKGDFATDESNTSFYHGDNGDKVVGFFTTGNNAWYADNNG 1398
Qy 1393 EALTGSORI-----NDQVRF-TREGQVKGDIYD 1422
Db 1399 NLVKGFOEIDGKWHFDEVGTGQAKGAALVNGQOLYFDVDSGIVQVKGDFVTD 1450

RESULT 13
GTF1_STRDO
ID GTF1_STRDO STANDARD; PRT; 1597 AA.
AC F11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfi;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX STRAIN=MFE28;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a Glucosyltransferase gene from Streptococcus
sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278 (1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
play a key role in the development of the dental plaque because of
their ability to adhere to smooth surfaces and mediate the
aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 19 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC
CC EMBL; M13931; AAC3063.1; -; Genomic DNA.
DR InterPro; IPR002479; Cell_wall_bd_put.
DR InterPro; IPR003318; Glyco_hydro_70.
```

```
DR Pfam; PF01473; CW binding 1; 4.
DR Pfam; PF02324; Glyco hydro 70; 1.
KW Dental caries; Glucosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1597 Glucosyltransferase-I.
FT REPEAT 1099 1132 A repeat.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1227 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.
FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 AC repeat.
FT REPEAT 1519 1568 AC repeat.
FT REPEAT 1582 1597 A repeat (incomplete).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1-25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MW; B9E86A20086798E CRC64;
```

```
Query Match 48.1%; Score 3587.5; DB 1; Length 1597;
Best Local Similarity 48.5%; Pred. No. 1e-142;
Matches 733; Conservative 236; Mismatches 400; Indels 141; Gaps 31;

Qy 1 METKRYKHKVKKHWTVAASGLITGLTTLGSSV-SAETEQQTSDKVVTKSEDDKA 59
Db 1 MEKNERFKHKVKKRWVTVSASA--TMLASALGASVASADTE-----TVSEDSNQA 50

Qy 60 AESSOTDAPKTKQATEQTEQQAQSOANVADTSTITKTPSQNIITQANSDDKTVTNKS 119
Db 51 VLTADQT-----TTNQDTEQ-----SVAATATS-----EOSADAATDQASATDQAS 94

Qy 120 EBAQTSEERTKQSEEAQTASS-----QALTQAKAELTKQRTAAQEN 162
Db 95 AAEQTQGTASTDTAAQTITANEAKWVTENENQVFTDEMLAEAKNVAT-----AESN 148

Qy 163 KNPDLAALPNVKQIDGKYIYIGSGQPKKNFALTVNNKVLFPDKNTGALTDTSQYQF-K 221
Db 149 SIPDLAKMSNVKQVDGKYIYDQGNVKKNFVSGEIKIYFDE-TGAYKDTSKVEADK 207

Qy 222 QG--LTKLNDYTPHQIVNPENTSLTIDNVYATDSWYRPKDLKNGKWTWASSDDL 279
Db 208 SGSDISKEETTFAANNRAYSTSAENFEADNLTADSWYRPKSLKDGKWTWESSKDDFR 267

Qy 280 PLLMGSWPKQTOIAYLYNMYNQGLTGENYTADSSQESLNLAQVTVQVKIETKISOTOQ 339
Db 268 PLLMAWVPDTEKRYVYVYVNVKV-VGIDKTYTAETSOADLTAAELVQARIEQKITTEQN 326

Qy 340 TQWLRIIDINSFVKTPQNMNSQTESPTSAGEKHOGGALLYSN-SDKTAYANSDYRLN 398
Db 327 TKWLREALSAFVKTPQNMNGSEKPYD---DHLQNGALKFDNQSDLTPTDQTSNYLLNR 382

Qy 399 TPTSQTG-----KPKYFEDNSSGGYDFLLANDIDNSNPVVOAQLNHLVLMYGSIVAND 454
Db 383 TPTNQTSGLDSRFTYNANDPLGGYBELLANDVNSNPVVOAQLNHLVLLNFGTIYAKD 442

Qy 455 PEANFDGVRVDAVDNVDNADLLQIASDYLKAHYGVDSKSEKNAIHLISILAWSNDNDPQYNK 514
Db 443 ADANFDSIRVDAVDNVDNADLLQISSDYLKAAVGIKNNKNNVNHVSIWAWSNDNDTPYLH 502

Qy 515 DTGKAQLPDKNLRLSLLYALTRPLEKQASNKNEIRSGLEPVITNSLNNRSAGEKNSERM 574
Db 503 DGDGDLNMDNKFRLSMLWSLAKPLDK-----RSLGNPLIHNSLVDRVDDREVEIV 554

Qy 575 ANYIIFRAHDSVQVVIKIIKAOINPKYDGLTFTLDELKQAFKYNEDMRQAKKYTOS 634
Db 555 PSYSFARHDSVQVQILIRDIKAEINPNAFGYSFTQDEIDQAFKYNEDLKKTKKYTHY 614

Qy 635 NIPTAYALMLNKSQITRLYYGDMYSDGQVYATKSPYYDAJDTLLKARIKYAAGQDMK 694
Db 615 NVPLSYTLLLTNKGSIPIVYIGDMFTDDGQYVANKVTYDAIESLLKARKMKVAGQAMQ 674

Qy 695 ITYVEGDKSHMDWDTGVLTSVRYGTGANEATDQGEATKTQGMVITNSNPSLKNQND 754
```

Db 675 -NYQINGE-----ILTSVRYGKALKQSKDATTTRTSVGVVMGNQNFSLD-GK 724
QY 755 KVVVNMGAHQOEYRPLLLTKDGLTSYTSDAAK--SLYRKTNDKGBLVDPASDIQY 812
Db 725 VVALNMGAHQOEYRALVSTKDGVTATYATDADASKAGLVKRTDENGILYFLINDDLKGV 784
QY 813 LNPQVSGYLAVVPVCGASDNQDVRVAASKANATQGVYESSALDSQLIYEGFSNPDV 872
Db 785 ANPQVSGFLQVVPVGAADQDILVAASDTADTGKSLHQDAMDSRVNMFEGFSNFQSPA 844
QY 873 TKDSYTNKIAQNVOLFSGWTSFEMAPQVYSSSDGSLDSIIQNGAFEDRDYDLAMS 932
Db 845 TKEEYTNVIANNVDPKFSWGLTFEMAPQVYSSDTGQFLDSVLQNGVAFTRDYDLGHS 904
QY 933 KNNKYSQODMINAVKALHKSIOVIADWPQIYNLPKEVVTATRVNDYGEYRKDSI 992
Db 905 KANKYGTADQLKAIKALHAKGLKVNADWPQMYTFPKQEVVTVTRTDKFKPIAGS01 964
QY 993 KNTLYAANTKSGKDYQAKYGAFSELAAKYPSIFNRTOISNGKKIDPSEKITAWKAKY 1052
Db 965 NNSLYVTDTKSGDDYQAKYGAFSELAKEPELFTKKQISTGQAIKQVSVKIKQWSAKY 1024
QY 1053 FNGTNILGRGVYLVKDNASDKYFELKGNQTLPLKQNTKEASTGPNVNDGNGMTF-YSTS 1111
Db 1025 FNGSNILGRGADYVLSQASNKLYNVDKDLFLPKTLTLLQGVVSGIRFDGTGVVNSST 1084
QY 1112 GVOAKNSFVODAKGNWYFDNNHMYGLQQLNGEVOYFLSNGVQLRESFLENADSKNY 1171
Db 1085 GERVTSDFITEA-GNLYYFGQDGYMTGAQNIKGSNYFLANGAALRNTVYTDQACQNH 1143
QY 1172 FGHGKNRYNGYYSFNDKSWYFVDSAGYMAVGLKTINGTQYFDDGYQVKGAWITGSD 1231
Db 1144 YGNDGKRYNGYQOQFNGDS-WRYF-RNGVWALGLTTVDGHVQVFDKQGVQAKDIIIVTRD 1201
QY 1232 GKRYFDDGSGNMAVNRFANDKNGDWYLYNSDGIALVGVQTNGKTYTFQDGKQIKGKI 1291
Db 1202 GKRYFDDQNGNAVTVTFVADKTGHWYLYGKQGVATGATVKGQHLYPEANGQVKGDF 1261
QY 1292 IT-DNGKLYFLANSSELARNIPATDSQNNWYFSGDGVAVTGSQTIAGKLYFASDGKQ 1350
Db 1262 VTAKDGKLYFDVDSGDMWNTFIEDKAGNWFYLGKDGAAVTGAQTIKGKLYFFKANGQ 1321
QY 1351 VKGSFV-TYNGKVHYHADSGELQVNR-----FEAD-----KQGN- 1384
Db 1322 VKGDIVKDADGKIRYYDAQTGEQVFNKSVSVNGKTYFSGDGTGTAQNPAGKQTFKQDGS 1381
QY 1385 -----WYILDSNGEALTGSRINDQRVFFTRGKQVKGQD-V 1419
Db 1382 VLRFYNLEGQYVSGSGWYETAHEWYVKS-GKVLGTGAQTIGNQRYVFXDNGHQVKGQLV 1440
QY 1420 AYDERLLVY 1429
Db 1441 TGNDGKLRVY 1450

RESULT 14

Q59983 9STRE PRELIMINARY; PRT; 1590 AA.
AC Q59983;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfi;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=QM2176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;

RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus";
RN DNA Seq. 4:19-27(1993).
[2]
RX PROTEIN SEQUENCE.
MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases";
J. Biol. Chem. 266:8916-8922(1991).
RL EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; 1HCX.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 47.9%; Score 3575; DB 2; Length 1590;
Best Local Similarity 48.4%; Pred. No. 3.4e-142;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

QY 1 METKERYKMHKVKHVTVAASGLITLGTTLGSSVSAETEQDTSKVVTKGSEDDKAA 60
Db 1 MEKNVRFKMKHVKRWILSVASA--TWLASALGASVAS-----ADDTTAS 44
QY 61 SESSOTDAPKTKQAQTEQTAQSANVADTSTSIKTPSQNITQAANSDDKTVNTKSE 120
Db 45 DDSNQA---VVTGDTQTNNAQTDQTSIAATATS-EQSASTDAATDQASAAEQGTAST 100
QY 121 EAQTSERTKQSEEAQ-----TTASSQALQAQKAEITKQRTAAQENKNPVDLAAIPVKQI 177
Db 101 D--TAAQTITNANEAKWPTENEGFTDEMLAEAKNVAT-AESDIPSDLAKMSNVKQV 157
QY 178 DGKYYYIGSDGPKPKNFALTNNKVLVFPKNTGALTDTTSQYQFKGLTKTLNND---YTPH 234
Db 158 DGKYYYDQDGNVKNFVSGVKIYYPDE-TGAYKDTSKVDADKSSSAVSQATIFAAN 216
QY 235 NOIVNFENTSLETDIINYVTADSWYRPFKQILKNGKTTWTSSESRLRPLLSMWPDKQTQIA 294
Db 217 NRAYSTSAKNFEAVDNYLTADSWYRPKSILKDGKTWTESGKDDFRPLLMWMPDPTETKRN 276
QY 295 YLYNNQQGLGTGENYTDSSQESINLAAQTVOVKIETKISQTOQTOWLRDINSFVKTK 354
Db 277 YVYNNKV-VGIDKTYTAETSOADLTAAAEVQARIEQKITSENNTKWLREAIASAFVKTK 335
QY 355 PNWNSQTSQTSAGEKQHLQGGALLYSN--SDKTAVANSDYRLNRTPTSTQGT---KPKY 409
Db 336 PWNGESEKPYD-----DHLQNGALLFDNQTDITPTDQSNRYRLNRTPTTQGTSLDSRFTY 391
QY 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNWLHLYLNYSIVANDPEANFDGVRVDAVDN 469
Db 392 NPNDPLGGYDFLLANDVDNSNPVQAEQLNWLHLYLNFGSIYANDADANFDSIRVDAVDN 451
QY 470 VNADLIQTASDYLKAHYGVKSEKNAIHLHSILEAWSNDNDPQYNKDTKAQLPIDNKLRL 529
Db 452 VDADLLQISSDYLKAAYGIDKNNKANNHVSIVEAWSNDNTPYLHDDGNLKNMMDKFR 511
QY 530 SLLYALTREPKDASKNBEIRSGLEPVTINLSNNSAEGKNSERMANYIFIRAHDEVOT 589
Db 512 SMLWLAKEPLDK-----RSLGPLIHNSIVDREVDRVETVPSYSFARAHDEVQD 563
QY 590 VIAIKIAQINPKTDGLTFLDELKQAFKIYNEDMRQAKKQYTSQNIPTAYALMLSNKDS 649
Db 564 IIRDIKAEINPNSFGYSFTQEBIEQAFKIYNEDLKKTKDKYTHYNVPLSYTILLTNKGS 623

Db 624 IPRVYDMFTDDGQYMANVTNYDAIESULLKARMKYVSGQAMQ-NYQIGNGE----- 676

QY 710 TGVLTSVRYGTGANEATDQGSBATKTQGMVITSSNPSLKLNONDKVIVNMGAAHKNQY 769

Db 677 --ILTSVRYGKALKQSDGDAATTRTSGVGVVGNQPNFSLD-GKVVALNMGAAHANQY 733

QY 770 RELLUTTKDGLTSYTSDDAAK--SLYRKTNDKGELVFDASDIQGYLNPOVSGYLAVWPEV 827

Db 734 RALMVSTKDGVAITYADADASKAGLVKRTDENGILYFLNDDLKGVANPQVSGFLOVWPEV 793

QY 828 GASDNODVRAASNAKANATQVYESSALDSOLIYEGFSNFQDFVTKDSDIYTNKKIAQNV 887

Db 794 GAADDQDIRVAASDTASTDGLSHQDAANDSRVMEGFSNFQSFATKEEYTNVVIANNV 853

QY 888 QLFKSWGVTSEFAPQYVSSDEGSFLDSIQNGYAFEDRYDLAMSKNNKYGSQQDMINAV 947

Db 854 DKFVSWGITDFEMAPQYVSSDTGQFLDSVIQNGYAFEDRYDLGMSKANKYGTADQLVKAI 913

QY 948 KALHKSQIOVIADWVDDQIYNLPKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSKNGK 1007

Db 914 KALHAKGLKVMADWVDDQMYTTPKQEVVTVTRTDKFKPIAGSQINHSLYVTDTKSSGDD 973

QY 1008 YOAKYGGAFLSALAAKYPSIFNRKTOISNGKKIDPSEKITAWKAKYFNGTNILGRGVGYVL 1067

Db 974 YOAKYGGAFLDELKEYPELFTYKKQISTQQAIDPSVKIKQMSAKYFNGSNILGRGADYVL 1033

QY 1068 KDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDGNGMTFYST-SCYQAKNSFVQDAKGN 1126

Db 1034 SDQVSNKYFNVASDRTLFLPSSLGKVVESGIRYDGKGYIYNSSATGDQVKASFITEA-GN 1092

QY 1127 WYFFDNNGHMVYGLQOLNGEVQYFLSNGVQLRESFLENADGSKNYFCHLGNRYSNGYYSF 1186

Db 1093 LYFFGKDGMYMTGAQTINGANYFFLENGTALRNTIYTDAGNSHYIYANDGKRYENGYYQF 1152

QY 1187 DNDSKWRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKRYFDDGSGNMAV 1246

Db 1153 GND--WRYF-KDGNMAGVITTVDGNVQYFDKGQVQAKDKLIIVTRDGKVYFDQHGNVAT 1209

QY 1247 NRPANDKNGDWYLNSDGIALVGVQTINGKTYTFQDQKQIKGKIITDN-GKLKYFLANS 1305

Db 1210 NTFIADKTGHWTYLGKDGVAVTGAQTVGKQKLYFEANGQVKGDFVTSHEGKLYFYDVDS 1269

QY 1306 GELARNIPATDSQNNWYFSGDGVAVTGSQTAGKLYFASDGKQVKGsfVT-YNGKVHY 1364

Db 1270 GDMWTDFTIEDKAGNMFYLGKGAAVSGAQTIRGQKLYFKAYGQVKGDIVKGTGDKIRY 1329

QY 1365 YHADSGELQVNRFEADKDG----- 1383

Db 1330 YDAKSGEVENKTVKKAADGKTYVIGNNGVAVDPSVVVGQTFKDGASGALRPFYNLKGQLVVG 1389

QY 1384 -----NWYYLDSNGEALTSQRINDQRFVFTREGKQVKGDV 1419

Db 1390 SGWYETANHDWVYIQS-GKALTGEQTINGQHLVFKEDGHQVKGQL 1433

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 158.264 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-37

Perfect score: 8349

Sequence: 1 MEKNVRFKMHVKRWVTLT.....VYQYYFNGDGARIYRGNW 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 segs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*
9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8349	100.0	1590	7	Add93657 Streptoco
2	8349	100.0	1590	9	Adx37280 Streptoco
3	8256	98.9	1592	2	Aar32925 Glucosylc
4	4672.5	56.0	1476	5	Aau79284 Streptoco
5	4580	54.9	1475	5	Aau98027 S. mutans
6	4580	54.9	1475	7	Add93654 Streptoco
7	4580	54.9	1475	9	Adx37277 Streptoco
8	4579	54.8	1475	5	Aau98030 S. mutans
9	4576	54.8	1475	5	Aau98040 S. mutans
10	4575	54.8	1475	5	Aau98031 S. mutans
11	4574	54.8	1475	5	Aau98033 S. mutans
12	4573	54.8	1475	5	Aau98032 S. mutans
13	4568	54.7	1475	5	Aau98035 S. mutans
14	4568	54.7	1475	5	Aau98034 S. mutans
15	4566	54.7	1475	5	Aau98036 S. mutans
16	4560	54.6	1475	5	Aau98037 S. mutans
17	4553	54.5	1475	5	Aau98039 S. mutans
18	4550	54.5	1475	5	Aau98038 S. mutans
19	4214.5	50.5	1375	5	Aau98028 S. mutans
20	4214.5	50.5	1375	5	Aau79288 Streptoco
21	4214.5	50.5	1375	7	Add93655 Streptoco
22	4214.5	50.5	1375	9	Adx37278 Streptoco
23	3587	43.0	1430	5	Aau98044 S. mutans
24	3583	42.9	1430	5	Aau98045 S. mutans

25	3582	42.9	1430	5	AAU98041	S. mutans
26	3580	42.9	1430	5	AAU98043	S. mutans
27	3578	42.9	1430	5	AAU98042	S. mutans
28	3575	42.8	1430	5	AAU98029	S. mutans
29	3575	42.8	1430	7	ADD93656	Streptoco
30	3575	42.8	1430	9	ADX37279	Streptoco
31	3525	42.2	1499	7	ADC54806	Protein S
32	3502.5	42.0	1554	7	ADD93658	Streptoco
33	3502.5	42.0	1554	7	ADX37281	Streptoco
34	3451.5	41.3	1017	5	AAU79285	Streptoco
35	3325.5	39.8	1518	7	ADD93660	Streptoco
36	3325.5	39.8	1518	9	ADX37283	Streptoco
37	3282.5	39.3	2835	5	ABR98574	Dextran 8
38	3282.5	39.3	2835	6	ABR55594	Amino aci
39	3245	38.9	1577	2	AAR91047	Alpha-D-g
40	3087.5	37.0	1527	5	AAU80055	Leuconost
41	3085.5	37.0	1527	7	ADC54807	Leuconost
42	3053	36.6	1477	9	ADY72696	Mutant de
43	3052	36.6	1477	9	ADY72733	Mutant de
44	3049	36.5	1477	9	ADY72732	Mutant de
45	3031	36.3	1365	7	ADD93659	Streptoco

ALIGNMENTS

RESULT 1
ADD93657
ID ADD93657 standard; protein; 1590 AA.
XX
AC ADD93657;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus sobrinus glucosyltransferase-I.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PE 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
(FORS-) FORSYTH INST.
PA Smith DJ, Taubman MA;
PI
DR WPI; 2003-845091/78.
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 16; Page 14; 49pp; English.
XX
CC The present sequence is the protein sequence of Streptococcus sobrinus
CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutans glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Dieptopic or
CC multi-epitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.

OS Streptococcus sobrinus.
XX JP05023188-A.
PN XX
XX XX
PD 02-FEB-1993.
XX XX
XX 25-JUL-1991; 91JP-00186592.
XX XX
XX 25-JUL-1991; 91JP-00186592.
XX XX
XX (KATO/) KATO K.
PA (FUKU/) FUKU I.
XX XX
XX WPI; 1993-079449/10.
DR N-PSDB; AAQ37760.
XX XX
XX DNA sequence glucosyl transferase-I - comprises Streptococcus sobrinus
PT DNA sequence with at least one nucleotide added or deleted.
XX XX
XX Claim 13; Page 15; 29pp; Japanese.
XX XX
XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes
CC glucosyl transferase-I (and mutants). The DNA was obtd. by treating S.
CC sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially
CC digesting with SauAI and fractionating on agarose gel. The 3-5 kbp
CC fragment was ligated into pUC18 and E. coli JM109 transformed with it. A
CC GT-1 expressing clone was isolated and sequenced. The clone may be used
CC in the development of a drug for dental caries
XX XX
XX Sequence 1592 AA;
SQ

Query Match 98.9%; Score 8256; DB 2; Length 1592;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1574; Conservative 6; Mismatches 10; Indels 2; Gaps 1;

QY 1 MEKNVFPKHKVKKRWVTLVSASATMLASALGASVASADTDASDSDNQAVVTGDDTTN 60
DB 1 MEKNVFPKHKVKKRWVTLVSASATMLASALGASVASADTDASDSDNQAVVTGDDTTN 60

QY 61 QATDQTSIAATATSEOSASTDAATDQASAAEOTOGTTASTDTAAQTNTTANAEAKWVPTEN 120
DB 61 QATDQTSIAATATSEOSASTDAATDQASAAEOTOGTTASTDTAAQTNTTANAEAKWVPTEN 120

QY 121 ENQGFTDEMLAEAKNVAATESDPSDLAKMSNVKQVGGKYYYDDQGNVKNFAVSVGD 180
DB 121 ENQGFTDEMLAEAKNVAATESDPSDLAKMSNVKQVGGKYYYDDQGNVKNFAVSVGD 180

QY 181 KIYYPDETGAJKDTSKVDADKSSSAVSQNAATIPAAANNRAYSTSAKNFEAVDNVLTADSWY 240
DB 181 KIYYPDETGAJKDTSKVDADKSSSAVSQNAATIPAAANNRAYSTSAKNFEAVDNVLTADSWY 240

QY 241 RPKSILKDGKTWTESGKDPRLMAWPDTEKRNVYNNKVVGIDKTYTAETSQADL 300
DB 241 RPKSILKDGKTWTESGKDPRLMAWPDTEKRNVYNNKVVGIDKTYTAETSQADL 300

QY 301 TAAAEHVQARIEQKITSENNWKLRSAISAFVKTQPMNGESEKPYDDHLQNGALLFDNQ 360
DB 301 TAAAEHVQARIEQKITSENNWKLRSAISAFVKTQPMNGESEKPYDDHLQNGALLFDNQ 360

QY 361 TDLTPDQTSNYRLNRTPTNQTSGLSDSRFTYNPNDFLGGDYFLANDVDNSNPVQAEQL 420
DB 361 TDLTPDQTSNYRLNRTPTNQTSGLSDSRFTYNPNDFLGGDYFLANDVDNSNPVQAEQL 420

QY 421 NWLHYLLNFGSIIYANDADANFDSIRVDAVDNVDADLLQISSDYLLKAAYGIDKNNKXANNH 480
DB 421 NWLHYLLNFGSIIYANDADANFDSIRVDAVDNVDADLLQISSDYLLKAAYGIDKNNKXANNH 480

QY 481 VSIWEAWSNDPTPLYLHDDGDNLMNDKNPRLSMLSLAKPLDKRSGINPLIHNLSVDREV 540
DB 481 VSIWEAWSNDPTPLYLHDDGDNLMNDKNPRLSMLSLAKPLDKRSGINPLIHNLSVDREV 540

QY 541 DDREVTVPSSYFARAHSEVQDIIRDIIKAEINPNPSFGYFSTQBEIDQAFKIYNEDLKK 600

DB 541 DDREVTVPSSYFARAHSEVQDIIRDIIKAEINPNPSFGYFSTQBEIDQAFKIYNEDLKK 600
QY 601 TDKKYTHYVPLSYTLTLTNKGSIPRVYVYGDMTDDGQYMANKTVNYDAIESLLKARMKY 660
DB 601 SDKKYTHYVPLSYTLTLTNKGSIPRVYVYGDMTDDGQYMANKTVNYDAIESLLKARMKY 660
QY 661 VSGQAMQNYQINGEILTSVRYGKALKQSDKGATTRTSVGVVGMGNOPNFSLDGKV 720
DB 661 VAGGQAMQNYQINGEILTSVRYGKALKQSDKGATTRTSVGVVGMGNOPNFSLDGKV 720
QY 721 ALNMGAALHANOQYRALMVSTKDVATYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
DB 721 ALNMGAALHANOQYRALMVSTKDVATYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
QY 781 PQVSGFLQVWPVGAADDQDQIRVAASDTASTDGKSLHQAAMDSDRMVFGFNFQSFATK 840
DB 781 PQVSGFLQVWPVGAADDQDQIRVAASDTASTDGKSLHQAAMDSDRMVFGFNFQSFATK 840
QY 841 EEEYTNVVIANNVDKFSVSGITDFEMAPQYVVSSTDGQFLDSVIQNGYAFTRDYDLGMSKA 900
DB 841 EEEYTNVVIANNVDKFSVSGITDFEMAPQYVVSSTDGQFLDSVIQNGYAFTRDYDLGMSKA 900
QY 901 NKYGTADQLVKAIKALHAKGLKVMADWVPDQMTYFPKQEVVTVTRTDKFGKPIAGSQINH 960
DB 901 NKYGTADQLVKAIKALHAKGLKVMADWVPDQMTYFPKQEVVTVTRTDKFGKPIAGSQINH 960
QY 961 SLYVTDTKSSGDDYQAKYGGAFELBELKEKYPFLTKKQISTQQAIDPSPVKIKQWSAKYFN 1020
DB 961 SLYVTDTKSSGDDYQAKYGGAFELBELKEKYPFLTKKQISTQQAIDPSPVKIKQWSAKYFN 1020
QY 1021 GSNILGRGADYVLSDOVSNKYNFNVASDTLFLPSSLLGKVESGIRYDGHGYIYNSSATGD 1080
DB 1021 GSNILGRGADYVLSDOVSNKYNFNVASDTLFLPSSLLGKVESGIRYDGHGYIYNSSATGD 1080
QY 1081 QVKASFITEAGNLVYFGKDGYNVMTGAQTINGANYFFLENGTALRNTIYTDAGQNSHYAN 1140
DB 1081 QVKASFITEAGNLVYFGKDGYNVMTGAQTINGANYFFLENGTALRNTIYTDAGQNSHYAN 1140
QY 1141 DGKRY--ENGYQFGNDWRYFKDGNMAGVGLTTVDGNVQYFDKDGVOAKDKIIVTRDGKVR 1198
DB 1141 DGKRYENGYQFGNDWRYFKDGNMAGVGLTTVDGNVQYFDKDGVOAKDKIIVTRDGKVR 1200
QY 1199 YFDQNGHNAATFTADKTGHVYILGKGVAVTGAQTGKQKLYFEANGQOVKGFVTS 1258
DB 1201 YFDQNGHNAATFTADKTGHVYILGKGVAVTGAQTGKQKLYFEANGQOVKGFVTS 1260
QY 1259 EGKLYFYDVGSDGMDTDTFIEDKAGNFWYLGKDGAAVTCGAQTIRGOKLYFKANGQOVKGD 1318
DB 1261 EGKLYFYDVGSDGMDTDTFIEDKAGNFWYLGKDGAAVTCGAQTIRGOKLYFKANGQOVKGD 1320
QY 1319 IVKGTGDKIRYDDAKSAGEQVFNKTVKAADGKTYVIGNDGVAVDPSVVGQTFKDAAGALR 1378
DB 1321 IVKGTGDKIRYDDAKSAGEQVFNKTVKAADGKTYVIGNDGVAVDPSVVGQTFKDAAGALR 1380
QY 1379 FYNLKGQVLTGSGWYETANHDVYIYQSGKALYTGEOQTINGOHLYFKEDGHQVKGQVLTGD 1438
DB 1381 FYNLKGQVLTGSGWYETANHDVYIYQSGKALYTGEOQTINGOHLYFKEDGHQVKGQVLTGD 1440
QY 1439 GKRVYDANSQDAFNKSVTVNGKTYVFCNGDTGTAOTAGNPKGQTFKDGSDIRFYSMEGOL 1498
DB 1441 GKRVYDANSQDAFNKSVTVNGKTYVFCNGDTGTAOTAGNPKGQTFKDGSDIRFYSMEGOL 1500
QY 1499 VTGSGWYENAOQWLYVKNKGKVLTLGLQTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDE 1558
DB 1501 VTGSGWYENAOQWLYVKNKGKVLTLGLQTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDE 1560
QY 1559 NSGSMITNWKQFYGVQYGYFFGNDGARIYRGWN 1590
DB 1561 NSGSMITNWKQFYGVQYGYFFGNDGARIYRGWN 1592

AAU79284 standard; protein; 1476 AA.
AAU79284;
13-AUG-2002 (first entry)
Streptococcus mutans monoclonal antibody-related protein #1.
XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;
KW glucosyl transferase-B; immunotherapy.
XX Streptococcus mutans.
OS Streptococcus mutans.
XX JF2002114709-A.
XX 16-APR-2002.
XX 04-OCT-2000; 2000JP-00304889.
XX 04-OCT-2000; 2000JP-00304889.
PR (UYN1-) UNIV NIPPON.
XX WPI; 2002-448885/48.
XX Anti-carries agent composed of a monoclonal antibody against an inhibitory
PT enzyme against water insoluble glucan synthetase of glucosyl transferase-
PT B (GTF-B) of Streptococcus mutans.
XX Claim 3; Page 13-16; 28pp; Japanese.
XX The invention relates to a monoclonal antibody against dental caries and
CC an anti-carries agent composed of a monoclonal antibody produced by
CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)
CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
CC inhibitive activity against water insoluble glucan synthetase of glucosyl
CC transferase-B. The monoclonal antibody specifically inhibits water
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
CC transferase-B and is used in the immunotherapy of dental caries. This
CC sequence represents a Streptococcus mutans monoclonal antibody-related
CC protein
XX
SQ Sequence 1476 AA;
Query Match 56.0%; Score 4672.5; DB 5; Length 1476;
Best Local Similarity 57.1%; Pred. No. 1.7e-274;
Matches 907; Conservative 210; Mismatches 332; Indels 139; Gaps 16;
QY 1 MEKNVRFKMHKVRWVTLVSASATMLASALGASVASADTDTA-----SDDSNAQVVTGQ 56
DB 1 MDKVRVYKLRKVRWVTVSVASAVMTLTLTSLGLVKADSNEKSQISNDSNTSVTANE 60
QY 57 TTN--NQATDQTSIAATATSEQASASTDAATDQASAAEQGTGTTASTDTAAQTNNANEAK 114
DB 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKE-----VVSNPYTVTGETASNGEKL- 115
QY 115 WYPTENENOGFTDEMLAEA-----KNVATAESDSIPSD-----LAKMSNVKQVDGKY 161
DB 116 -----QNTTVDKTSAAANNISKQTEADTVIDDSSNAANLQLEKLPNVKEIDGKY 169
QY 162 YYDDQGNVKNFVSVGDKIYIFDETGAJKDTSKVDADKSSASVSONATIFAANNRAYS 221
DB 170 YYDDNNGKVRTNFTLADGKILHFDGTGAYTDTSIDTVNK--DIVTTRSNLYKKYNQVD 227
QY 222 TSAQNFPAVDNLTADSWTRPKSILKDGKTWTESGKDDRPRLMAWPDPTETKRNVTNM 281
DB 228 RSAQSFHVDHLYTAESRPRPKYILKDGKTWTQSTEKDFRPLMTWPPDQETQRQVNTM 287
QY 282 NKVVVGIDKTYTAETSQADLTAAELVQARIEOKITSENNTKWLREALSAFVKTPQWNGE 341
DB 288 NALGLINKYTDTSNQLUNIAAATQAKIEAKITFLKNQDWLRQTISAFVKTPQWNSD 347
QY 342 SEKPYDDHLQNGALLFDNQDLDLTPDQTSNYRLNRPNTQNTGSLDSRFTYFNPNDPLGGYD 401

DB 348 SEKPFDDHLQNGAVLYDNEGLTPYANSNYRIILNRTPTNQTGKKDPYV--ADNTIGYE 405
QY 402 FLLANDVDNSNPVVQAEOLNHLNFGSIYADADANFDSIRVDVNDVADLLOISS 461
DB 406 FLLANDVDNSNPVVQAEOLNHLNFGSIYADADANFDSIRVDVNDVADLLOIATG 465
QY 462 DYLKAAAGIDKNNKANNHVSIVAEASNDNTPYLHDDGDMNNDKFRLSMLWSLAKPL 521
DB 466 DYLKAAAGIHKNDKANDHLSILEASNDNTPYLHDDGDMNNDKFRLSMLWSLAKPL 525
QY 522 DKSGNLPLHNSLDREVDRVETVPSYSFAPAHSDSEVQDIIRDIKAEINPNPSFGYS 581
DB 526 NQSGMNPITNSLVNRTDDNAETAAVPSYSFIRAHSDSEVQDLIRDIKAEINPNVVGYS 585
QY 582 FTCEETEQAFKIYNEDELKKTCKKYTHYVNPVLSYTLTLTNKGSIPRVYVGMFTDDGQYMA 641
DB 586 FTMBEIKCAPEIYNKOLLATEKIKYTHYNTALSYALLTNKSYPRVYVGMFTDDGQYMA 645
QY 642 NKTWNYDAJESLLKARMKYVSGQAMQNTYOIGNGEILTSVRYGKGLKQSDKGDATRTS 701
DB 646 HKTINYEATETLLKARIKYVSGQAMRNQVGNSEITTSVRYGKGLKATDGTGRTTTS 705
QY 702 GUGVVMGNQPNFSLDGK--VVALNMGAAHANQEYRALMVSTKDGVTATYATDADASKAGLVK 760
DB 706 GVAVIEGNNPSLRKASDRVVMNGAAHKQAYRPLLLTTDNGIKAYHSQDEA--AGLVK 763
QY 761 RTDENGILYFLNDDLKGVANPQVSGFLOVWPVGAADDQDQIRVAASDASTDGSKLHODA 820
DB 764 YNDRGELIFTAADIKGYANPQVSGYLVGVVPGAADQDVRVAASTPSTDGSKVHONA 823
QY 821 AMDSRVWFSGFNFQSFATKEBEYTNVIANNVDFKFSWGITDFEMAPQVSVSTDGQFLD 880
DB 824 ALDSRVWFSGFNFQSFATKEBEYTNVIANNVDFKFSWGITDFEMAPQVSVSTDGQFLD 883
QY 881 SVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPQDQMTFFPKQEV 940
DB 884 SVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPQDQMTFFPKQEV 943
QY 941 VTVTRTDKPKIAGSQINHSILYVTTKSSGDDYQAKYGGAFLDEKKEYPFELTKKQIS 1000
DB 944 VTATRVKYGTPTVAGSQIKNTLYVVDGKSSGKQQAQYGGAFLEELQAKYFELFARKQIS 1003
QY 1001 TQQAIDPSVKIKOWSAKYFNGSNILGRGADYVLSDOVSNNKYFNVA--SDTLFELPSSILGK 1058
DB 1004 TGVPMWPSVKIKOWSAKYFNGSNILGRGADYVLSDOVSNNKYFNVA--SDTLFELPSSILGK 1063
QY 1059 VYESGIRYDGKGIYNSSATGQVKAFTFTEAGNLYYFGKDGVMVTGAQTINGANYPFLE 1118
DB 1064 DSQVGFSDYDGKGVYVYST--SGYQAKWTFISEGDKWTFYFDNNGVMVTGAQSGINGVNYFLS 1122
QY 1119 NGTALNTIYTDQGNSHYANDKRYENGYQOF--GNDWRYPKDGMNVAGLTIVDGNVQY 1177
DB 1123 NGLQLRALDLKNEDEGTIAYYGNDRRYENGYQFMSGVWRHFNNGEMSVGLTIVDGVQY 1182
QY 1178 FDKDGVQAKKILVTRTDGKRVYFDQHNAAATTFIADKTGHVYLGKGVATVGAQTVG 1237
DB 1183 FDMGYQAKGKFTTADGKIRYFDKOSGNWNRNRFTEBEGRWLYLGEDGAAVTSQITIN 1242
QY 1238 KQKLYFEANGQVKGDFVTSDEGKLYFYDVSDDMMWTDTFIEDKAGNWFYLGKDGAAVTC 1297
DB 1243 GQHLFYFRANGVQVKGFEVTDHGRISYYDNGSDQIRNRFVRNAQOQWYFDNNGYAVTC 1302
QY 1298 AOTIRGOKLYFKANGQVKGDI VKGTDGKIRYDAKSGSQVFNKTVKAADGKTYVTGNDG 1357
DB 1303 ARTINGQHLFYFRANGVQVKGFEVTDHGRISYYDNGSDQIRNRFVRNAQOQWYFDNNG 1362
QY 1358 VAVDPSVVKQTFKDSAGALRYNLKQLVTGSGWVETANHDWVYIQSCKALTGEOTING 1417
DB 1363 YAV-----TGARTING 1373
QY 1418 QHLYFKEDGHQVKGQVLTGTDGKRVYDANSGDQAFNKSVTVNGKTYFENDCTAOTAGN 1477

Db 1374 QHLYFRANGVQKGEFVDRHGRISYYDNGSGDQIRNRFV----- 1413
 Qy 1478 PKGQTFKGDSDIRFYSMEGQLVTGSGWYENAAQOWLYV- KNGKVLRLGLQTVGSQRVYFDE 1536
 Db 1414 -----RNAQOWFYFNDNGVAVTGARTINGQHLFYFA 1445
 Qy 1537 NGTQAKGKAVRTSDGKIRYFDENSGSMI 1564
 Db 1446 NGVQVKGFTVDRYGRISYYDANSGRV 1473

RESULT 5

AAU98027
 ID AAU98027 standard; protein; 1475 AA.

AC AAU98027; 1

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB.

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture.

XX Streptococcus mutans.

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

DR N-PSDB; ABK52938.

XX Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in paper

XX manufacture, comprises mutations in specific positions.

XX Disclosure; Page 21-25; 44pp; English.

CC The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from I48V, D457N, D567T,

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,

CC I448Y/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents GTFB
 XX
 SQ Sequence 1475 AA;

Query Match 54.9%; Score 4580; DB 5; Length 1475;

Best Local Similarity 56.4%; Pred. No. 7e-269;

Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;

Qy 1 MEKNVRFKMHKVKRWVTLVSASATWLASALCASVASADTDTA----SDDSNOAVVTGQ 56

Db 1 MDKRVYKLRKVKRWVTVSVASAVMTLTLSGGLVKADSNESKQISNDSNTSVVTANE 60

Qy 57 TTN--NOATDOTSIAATATSEQASASTDAATQASAAEQGTGTTASTDTAAQTNNANEAK 114

Db 61 ESNVITEATSKQEAASSQINHTVTTSSSSTSVVNPK-----VVSNPYTVGETASGEKL- 115

Qy 115 WYPTENEGFTDEMLAEA-----KNVATASDSIPSD-----LAKSNVVKQVDGKY 161

Db 116 -----QNGTTTVDKTTSEAAAANNISQGTTEADTVIDDSNAANLQLEKLPNKETDGY 169

Qy 162 YYDODGNVKNFAVSGDKIYVFDETGAYKDTSKVDADKSSSAVSONATIFAANNRAYS 221

Db 170 YYDNNGKVRTNFTLIADGKILHFDETGAYTDTSDTVNK--DIVTTRNLYKYNQVVD 227

Qy 222 TSAKNFEAVDNYLTADSWYRPKSILKDGKTWTESGKODFRPLLMAWPPDTETKRNYYNM 281

Db 228 RSAQSFEHVDHYLTAESWYRPKYILKDGKTWTQSTEKDFRLLMTWPPDOETQORQVNYM 287

Qy 282 NKVVGIDKTYTAETSQADLTAAAEVQARI EOKIITSENNTKWLREAI SAEVKTQPOWNGE 341

Db 288 NAQLGINKTYDDTSNQLQNLNIAAATQAKIEAKITTLKNTDWLRQTISAFVKTSANSD 347

Qy 342 SEKPYDDHLQNGALLFDNOTDLTPTQSNVRLNRPNTNQTGSLDSRFTYNNPNDPLGGYD 401

Db 348 SEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRLNRPNTNQTGKDPRT--ADNTIGGYE 405

Qy 402 FLLANDVDSNPNVQAEQLNLWLHLYLNFSGSIYANDADANFDSIRVDAVDNVDADLLQISS 461

Db 406 FLLANDVDSNPNVQAEQLNLHFLHFLMNFNFIYANDPDANFDSIRVDAVDNVDADLLQIAG 465

Qy 462 DYLKAAAGIDKKNKANNHVSIVEASNDTPLYLHDDGDNLMNMKNKFLMLWSLAKPL 521

Db 466 DYLKAAAGLHKNDKAANDHLSLEASNDTPLYLHDDGDNMNMKNKLSLFLSLAKPL 525

Qy 522 DKRSGINPLIHNSLVDRVDDREVEVTPSYSPARAHDSVQDIIRDI IKAENPNPSFGYS 581

Db 526 NQSGNPNLITNSLVNRTDDNDAETAAVPSYFIRAHDSVQDDIADI IKAENPNPVGYS 585

Qy 582 FTQEEITEQAFKINEDLKKTKKTHYNNVPLSYTLTLLTNKSGIPRYVYVYDMDTDDGOYMA 641

Db 586 FTMEIEIKAFIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYDMDTDDGOYMA 645

Qy 642 NKTVNDYDALESLLKARKKTVSGQAMQNTQIENGELLTTSRYGKGLKOSDKGDATRTS 701

Db 646 HKTINYEALETLLKARIKTVSGQAMRNQOVGNSEIITSVRYGKGLKATDTGDRTRTS 705

Qy 702 GGVVWNGPNPFLSDGK-VVALNMGAAHAAQRYALMVSTKQGVATYATDADASKAGLVK 760

Db 706 GVAIEGNNPBLKASDRVVVVMGAHKNQAYRPLLLTTDNGIKAYHSQEA--AGLVR 763
QY 761 RTDENGYYFLNDDLKGAVANPOVSGFLQVWPVGAADDQDIRVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGAVANPOVSGYLVWVPVGA--LTKFALRLARPHQOMASVHQ 820
QY 819 DAAMDSEVMEGSGNSFQSPATKEEYTNVVIANNVDKFSVWGITDFEMAPQYVSSDGGP 878
Db 821 NAALDSEVMEGSGNSFQSPATKEEYTNVVIANNVDKFAEWGVTDFEMAPQYVSSDGSF 880
QY 879 LDSVIQNGYAFTDRYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQWYFPKQ 938
Db 881 LDSVIQNGYAFTDRYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQWYFPK 940
QY 939 EVTVTRTRDKFGPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKYEPELFTKKQ 998
Db 941 EVTVTRTRDKFGPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKYEPELFTKKQ 1000
QY 999 ISTQOALDPSVKIKQSAKYPNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSL 1056
Db 1001 ISTQOALDPSVKIKQSAKYPNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSL 1060
QY 1057 GKVVESGIRVDGKGYLYNSSATGDQVKAFTIAGNLYYFGKGYVMTGAOTINGANYEP 1116
Db 1061 NQDSQVGFSDGKGYLYNST-SGYQAKNTFISGDKWYFPDNNGYMTGAQSGINGVNYF 1119
QY 1117 LENGTLRNTIYDAQNSHYANDGKRYENGQQF-GNDWRFKDGNMAVGLTIVDGNV 1175
Db 1120 LSNGLQLRALDKNEGDYVAYGNDGRYENGYYQFMSGWRHFNNGEMSVGLTIVDGV 1179
QY 1176 QYFDKQGVQAKIIVTRDGKRYVFOHNGNAATNTFIADKTHWYLGKGVAVTGAQT 1235
Db 1180 QYFDKQGVQAKIIVTRDGKRYVFOHNGNAATNTFIADKTHWYLGKGVAVTGAQT 1239
QY 1236 VQKQLYFFRANGQVKGDFVTSDEGLYFYDVSQDMWTDFTIEDKAGNWFYLGKGAAY 1295
Db 1240 INQHLIFRANGQVKGDFVTSDEGLYFYDVSQDMWTDFTIEDKAGNWFYLGKGAAY 1299
QY 1296 TGAQTIRGKLYFKANGQVKGDIKGTGKIRYDAKSGEQVFNKVAADGKTIVIGN 1355
Db 1300 TGARTINGQLLYFRANGQVKGDFVTSDEGLYFYDVSQDMWTDFTIEDKAGNWFYLGKGAAY 1359
QY 1356 DGAVDPSVVGKGTFKDASGALRYFNKLGQVGTSGWYETANHDWYIYOSKALTCGEQTI 1415
Db 1360 NGYAV-----TGARTI 1370
QY 1416 NGQHLIFKEDGQVKGDFVTSDEGLYFYDVSQDMWTDFTIEDKAGNWFYLGKGAAY 1475
Db 1371 NGQHLIFRANGQVKGDFVTSDEGLYFYDVSQDMWTDFTIEDKAGNWFYLGKGAAY 1475
QY 1476 GNPKGQTFKDGSDIRFYSMEGQVGTSGWYETANHDWYIYOSKALTCGEQTI 1534
Db 1413 -----RAGQGVFFDNGGYAVTGARTINGQLLYF 1442
QY 1535 DENGIOAKGAVRTSDGKIRYFDENSGSMI 1564
Db 1443 RANGQVKGDFVTSDEGLYFYDVSQDMWTDFTIEDKAGNWFYLGKGAAY 1475

RESULT 6

ADD93654
ID ADD93654 standard; protein; 1475 AA.

XX AC ADD93654;

XX DT 29-JAN-2004 (first entry)

XX DE Streptococcus mutans glucosyltransferase-B.

XX KW Glucosyltransferase; enzyme; vaccine; anticonvulsants; immunogen.

XX OS Streptococcus mutans.

XX

PN XX WO2003075845-A2.
XX PD 18-SEP-2003.
XX PF 07-MAR-2003; 2003WO-US0006962.
XX PR 07-MAR-2002; 2002US-0363209P.
XX PR 08-AUG-2002; 2002US-0402483P.
XX PA (FORS-) FORSYTH INST.
XX PI Smith DJ, Taubman MA;
XX PI WPI; 2003-845091/78.
XX PT Composition useful as vaccines for dental caries comprises a fragment of
XX PT a glucan binding protein-B binding to a major histocompatibility complex
XX PT class II protein.
XX PS Claim 16; Page 12-13; 49pp; English.
XX CC The present sequence is the protein sequence of Streptococcus mutans
XX CC glucosyltransferase-B (GTF-B). Peptide fragments of GTF-B, especially
XX CC from the catalytic domain of the polypeptide, can be used in immunogenic
XX CC compositions and subunit vaccines for dental caries. These compositions
XX CC comprise a major histocompatibility complex (MHC) class II protein-
XX CC binding peptide from S. mutans glucan binding protein-B (GbpB)
XX CC covalently linked with a peptide fragment of a streptococcal
XX CC glucosyltransferase. The compositions are used in a claimed method of
XX CC eliciting production of an antibody in a mammal. Diepitopic or
XX CC multi-epitopic polypeptides can be prepared synthetically or by
XX CC recombinant DNA technology. Antibodies raised against MHC class II
XX CC binding fragments of GbpB can be used in passive immunisation.
XX SQ Sequence 1475 AA;
Query Match 54.9%; Score 4580; DB 7; Length 1475;
Best Local Similarity 56.4%; Pred. No. 7e-269;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
QY 1 MEKVRFPKHKYKRWVTLVSASATWLASALGASVASADTDFA----SDSRQAVVTGDQ 56
DB 1 MDKVRYPKHKYKRWVTLVSASATWLASALGASVASADTDFA----SDSRQAVVTGDQ 56
QY 57 TTN--NOATDQTSIAATATSEQASASTDAATQASAAEQGTGTASTDTAQTITNANEAK 114
DB 61 ESNVTEATSKQEAASSQTNHTVTTSSSTSVVNPKE-----VVSNPYTVGTASNGEKL- 115
QY 115 WYPTENENQGFDEMLAE-----KNVATAESDSIPSD-----LAKMSNVKQVDGKY 161
DB 116 -----NQTTTVDKTSEAAANNISQOTTEADTDVDDSNAAALQLEKLPNVKEIDGKY 169
QY 162 YYYDQGNVKNFPAVSGDKIYYFDETGAYKDTSKVDADKSSASVSQNTATIFANNRAYS 221
DB 170 YYYDQGNVKNFPAVSGDKIYYFDETGAYKDTSKVDADKSSASVSQNTATIFANNRAYS 227
QY 222 TSARKFEAVDNYLTADSWVRPKSILKDGKWTWESKDDPRLMAMWPDTEKKNVNYM 281
DB 228 RQAQFEHVDHYLTAEWSVRPKSILKDGKWTWESKDDPRLMAMWPDTEKKNVNYM 287
QY 282 NKVVGIDKTYTAETSQADLTAAAEVLQARIEQKITSENNKWLREAIASAFVKTQPOWNGE 341
DB 288 NAQLGINKTYDDTSNQLQNLNIAAATIQAKIEKITTLKNYDNLRTQISAFVKTQPOWNGE 347
QY 342 SEKPYDDHLQNGALLFDNQTDLTPQTQSNRYLLNPTNTQTSLSRFRFPNPNPGLGYD 401
DB 348 SEKPYDDHLQNGALLFDNQTDLTPQTQSNRYLLNPTNTQTSLSRFRFPNPNPGLGYD 405
QY 402 FLANDVNSNPVQAEQLNMLHLYLLNFGSIYANDADANFDSIRVDADVNDADLQIS 461
DB 406 FLANDVNSNPVQAEQLNMLHLYLLNFGSIYANDADANFDSIRVDADVNDADLQIS 465
QY 462 DYLKAAAGIDKKNKNNHVSIVEAWSNDPTPLDGDGNDGNMNDKFNLSLWSLAKPL 521

```
Db 466 DYLKAAKGIHKNDKAAANDHLSILEAMSDNDTPYLHDDGDMNMDNKNLRLSLFLSLAKPL 525
Qy 522 DKRSGNLPLIHNLSLVREVDVREVPVSYSPARAHDSVQDIIRIIKAEINPNPFGYS 581
Db 526 NORSNGNPLITNSLVNRDNDNAATAAPSPSYSFTRAHDSVQDLIADIIKAEINPNVVGYS 585
Qy 582 FTQEEIEQAPKIVNEDLKTDKKYTHYNNPLSVTLALTNGKSI PRVYVYGMFTDDGQYNA 641
Db 586 FTWEEIKKAPFIYNNKOLLATEKKYTHYNTALSVALLLTNRKSSVPRVYVYGMFTDDGQYNA 645
Qy 642 NKTVNYDAIESLLKARMKVYSGGQAMQYQINGEILTSVRYGKGALKQSKDGADATRTS 701
Db 646 HKTINYEAETLLKARIKVYSGGQAMRNQVGNSEIITSVRYGKALKATDGTDRTRIS 705
Qy 702 GUGVNGNQNPSLDGK - VVALNMGAAHANQEYRALMVSTKDGVAFYATDADASKAGLVK 760
Db 706 GVAVIEGNNPSLRKASDRVVNMGAAHKNAQYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
Qy 761 RTDENGYLFLNDLKGVANPOVSGFLQVWVPVGAADDQDIRVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAADIIGYANPOVSGYLGWVPVGA--LIRKFAURLARPHQQMASVHQ 820
Qy 819 DAAMDNRVMEFSGFSQATKEEYTNVVIANNVDKFSWGITDFEMAPQYVSSTDGQF 878
Db 821 NAALDSRVMEFSGFSNFQAATKEEYTNVVIANNVDKFAEWGVTDFEMAPQYVSSTDGSP 880
Qy 879 LDSVIQNGYAFTDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMTFPKQ 938
Db 881 LDSVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSKGIKVMADWVPDQMYAFPEK 940
Qy 939 EVVTVTRDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDELKEKYPELTKKQ 998
Db 941 EVVTVATRVDKYGFPAVAGSQIKNTLYYVDGKSGQQAQYGGAFBELOAKYPELFARKQ 1000
Qy 999 ISTGQAIQPSVKIKQWAKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSL 1056
Db 1001 ISTGVPMDFSVKIKQWAKYFNGTNILGRGAGYVLDQATNYFNI SDNKEINFLPKILL 1060
Qy 1057 GKVVESGIRYDGKGYIYNGSATGDQVKASFITBAGNLVYFGKQVYMTGAQTINGANYFF 1116
Db 1061 NQDSQVGFSDYDGKGYVYIST -SGYQAKNTFISEGDKWYTFDNGVYMTGAQSGINGVNYFF 1119
Qy 1117 LENGTLARNTIYTDAGNSHYANDGKRYENGYQOF -GNDWRYFKDGNMVAGLTIVDGNV 1175
Db 1120 LSNGLQRLDAILKNEBDGTVAYNGDGRYENGYQFMSGVWRHFNNGEMSVGLTVIDGQV 1179
Qy 1176 QYFDKQGVQAKDIIIVTRDGKVRYFQHNAGNAATNTFIADKTGHYYLKGDVAVTGAQT 1235
Db 1180 QYFDEMGYQAKGFVTTADGKIRYFDQSGNMYRNFIEENEGKWLVLGEDGAATVGSQT 1239
Qy 1236 VGKQKLYFEANGQOVKGDVFTSDEGLKYFYVDVDSGDMWTDFTFIEDKAGNWFYLGKDGAAV 1295
Db 1240 INGCHLYFRANGVQVKGFEVTDHGRISYYDNGSGDQIRNRFVRNAQGGQVFPDNGYAV 1299
Qy 1296 TGAOTIRGQKLYFKANGQOVKGDIVKGTGKIRYDQKIRYDAKSGEQVFNKTVKADGTYVIGN 1355
Db 1300 TGARTINGQLLYFRANGVQVKGFEVTDYGRISYYDNGSGDQIRNRFVRNAQGGQVFPDN 1359
Qy 1356 DGVAVDPSVVKGQTFKXDGALRYNLKGLVTGSGWYETANHDWYVIOGSKALTGEOI 1415
Db 1360 NGYAV-----TGARTI 1370
Qy 1416 NGQHLVFKEDGHQVKQLVGTGTDGKRYVYDANSQDQAFNKSVTNKGTYVFGNDGTAQTA 1475
Db 1371 NGQHLVFRANGVQVKGFEVTDHGRISYYDNGSGDQIRNRFV----- 1412
Qy 1476 GNPQGTQFKDGSIRFYSMEQLVTGSGWYENAGQWLIV -KNGKVLTLGLQTVGSRVYF 1534
Db 1413 -----RNAQGGQVFPDNGYAVTGARTINGOHLV 1442
Qy 1535 DENGIOAKGAVRTSDGKIRYFDNDSGSMI 1564
Db 1535 DENGIOAKGAVRTSDGKIRYFDNDSGSMI 1564
```

```
Db 1443 RANGVQVKGFEVTDYGRISYYDANSGERV 1472
RESULT 7
ADX37277
ID ADX37277 standard; protein; 1475 AA.
XX AC ADX37277;
XX DT 21-APR-2005 (first entry)
XX Streptococcus mutant glucan binding protein B variant #6.
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX Streptococcus mutans.
XX US2005031633-A1.
XX 10-FEB-2005.
XX 09-MAR-2004; 2004US-00797821.
XX 13-APR-1998; 98US-0081550P.
XX 08-JAN-1999; 98US-0115142P.
XX 12-APR-1999; 99US-00290049.
XX 07-MAR-2002; 2002US-0363209P.
XX 08-AUG-2002; 2002US-0402483P.
XX 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
XX (TAUB/) TAUBMAN M A.
XX Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX Claim 7; SEQ ID NO 34; 73pp; English.
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX SQ Sequence 1475 AA;
Query Match 54.9%; Score 4580; DB 9; Length 1475;
Best Local Similarity 56.4%; Pred. No. 7e-269;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
Qy 1 MEKNVRFKQHKYKRWVTLVSASATMLASALGASVASADTDTA----SDDSNQAVVTGDQ 56
Db 1 MDKVRKYLKRVKRWVTVSVASAVMTLTLSGGLVKADSNESKSQISNDSNTSVVTANE 60
Qy 57 TTN--NQATDQTSIAATATSEQASDADTAADQASAEQTCGTTASTDTAAQTNNANEAK 114
Db 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKE----VVSNPYTVGETASNGEKL- 115
Qy 115 WYPTENENGFTDEMLAEA-----KNVATASDSIPSD-----LAKMSNVKQVVDGKY 161
Db 116 -----QNTTVDKTSSEAAANNISKQTEADTDVDDNSNAANLQILEKLPNVKEIDGKY 169
Qy 162 YYYDQNGNKKNFVSVGDKIYYFDBTGAYKOTSKVDADKSSSAVSQNTAFIPANNRAYS 221
Db 170 YYYDNGKVRTNFTLIADGKILHFBETGAYTDTSIDTVNK--DIVTTRSNNLYKKYQVVD 227
```


Qy	222	TSKQNFVAVNYLTADSWYRPSKSLKDGKWTWESGDKDPRPLLMAMWPDPTETKRYNYNM	281
Dd	228	RSQSFVHDVHLYTAESWYRPKYLKDGKWTWQSTEDKFRPLMTWMPDQETQRYNYM	287
Qy	282	NKVVGIDKTYTAETSQADLTAAAEVLQARIEOKITSENNYKWLBEALSAFVKTPQWNGE	341
Dd	288	NAQLGINKTYDDTSNQLNLAAATIQAKIEAKITLLKNTDMLRQTTSAFVKTSQSAWNSD	347
Qy	342	SEKPYDHLQNGALLFQNTDLTPDTQTSYRLNLRPTNTQSGSDSRFTYNNPDPLGGYD	401
Dd	348	SEKPFDDHLQNGAVLYDNEKLIPTYANSYRILNRTFTNTGKDPYAT--ADNTIGGYE	405
Qy	402	FLIANDVDSNPVQABQLNLHLYLNFSGSIYANDADANFDSIRVDVNDVADLQIASS	461
Dd	406	FLIANDVDSNPVQABQLNLHLYLNFSGSIYANDADANFDSIRVDVNDVADLQIAG	465
Qy	462	DYLKAAAGVGDKNKNANHHVSIYVANSNDNTPYLHDDGDMNMDNKNFRLSMLSLAKPL	521
Dd	466	DYLKAAAGVHKNDAKNDHLSLILEANSNDNTPYLHDDGDMNMDNKNLRLSLFLSLAKPL	525
Qy	522	DKRSGNLPLTHNSLVREVDREVTVPYSFARAHDSVQDIIRDIKAEINPNSFGYS	581
Dd	526	NQRSGNPLITNSLVNRTDNEATAVPSISFIRAHDSVQDIIRDIKAEINPNSVGYIS	585
Qy	582	FTQEEIEQAFKINEDLKTKKTYTHYVPLSYLLTNKSGIPRVYIGDMFTDDGQYMA	641
Dd	586	FTMEIEKKAPEIYNKOLLATEKKYTHYNTALSYALLTNKSSVPRVYIGDMFTDDGQYMA	645
Qy	642	NKTAVNYDAESLAKARMKYSGQAMQNYQIGNEILTSVRYKGKALKQSKDGADATRTS	701
Dd	646	HKTINYEATLETLKARIKYSGQAMRNOQVNSEIITSVRYKGKALKATDGTGDRTRTS	705
Qy	702	GUGVWGNQPNFSLDGK--VVALNNGAAHANOEVRLMWSTKDGCVATVATDADAKAGLVK	760
Dd	706	GVAVIEGNNPSELKASDRVNVNNGAAHAKNOAVRPLLLTTDNGIKAYHSDQEA--AGLVR	763
Qy	761	RTDENGVLPLNDLKGAVNPQVSGFLQWVVPVGAADDODIRVAASDTASTDGK--SLHQ	818
Dd	764	YTNDRGELIFTAADIKGAVNPQVSGVLGVWVPVGA--LTKMFALRLARPHQOMASVHQ	820
Qy	819	DAAMDSRVNFEFSNFQSFATKEEYTNVVIANNVDFVSWGIGITDFEMAPQYVSSDGGF	878
Dd	821	NAALDSRVNFEFSNFQAFATKEEYTNVVIANNVDFVSWGIGITDFEMAPQYVSSDGGF	880
Qy	879	LDSVIONGVAFTDRYDLGHSKANKYGTADOLVKAIKALHAKGKVMWADVPDQMYTFPKQ	938
Dd	881	LDSVIONGVAFTDRYDLGHSKANKYGTADOLVKAIKALHAKGKVMWADVPDQMYTFPEK	940
Qy	939	EVVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKYPELFTKKQ	998
Dd	941	EVVTVTRDKYGTFTVAGSQIKNTLYVVDGKSSGKQOQAKYGGAFLELOAKYPELFPARKQ	1000
Qy	999	ISTGQADISVKIKQWSAKYFNGSNILGRADYVLSQVSNKYFNVA--SDTLFLPSSLL	1056
Dd	1001	ISTGVPMDFSVKIKQWSAKYFNGSNILGRADYVLSQVSNKYFNVA--SDTLFLPSSLL	1060
Qy	1057	GKVVESGIRYDGKGLYNSATGDQVKSFLTEAGNLVYFGKQVMTGAQTNGANYFF	1116
Dd	1061	NQDSQVGFSDGKGLYNSATGDQVKSFLTEAGNLVYFGKQVMTGAQTNGANYFF	1119
Qy	1117	LENGTALRNTIYTDAGNSHYANDGKRYENGVOQF--GNDWRYFKDGNMVAVGLTVDGNY	1175
Dd	1120	LSNGLQLRALILKNEGTAYYCNDRRYENGVOQFSGVGRHFNNGEMSVGLTVDGQV	1179
Qy	1176	QYFDKQGVQAKKIIIVTRDGKRYFQHNNGNAATNTFIADKTHWYLYLKGDKGVAVTGAQT	1235
Dd	1180	QYFDEMGYQAKGFVTTADKIRYFDKQSGNRYNRNFIENEKWLILYLGEDGAAVTGSQT	1239
Qy	1236	VGKQKLYFEANGQVKGDFVTSDEGLYFYDVDSGDWMTDTFTIEDKAGHWFVGLKDGAAV	1295
Dd	1240	INGQHLHYFRANGVQVKGFEVTDHHRISYIDGNSGSDQIRNRFVRNAQGMFVFDNNGYAV	1299
Qy	1296	TGAQTIRGOKLYFRANGQVKGDIVKGTGDKIRIYDAKSGEQFENKTVKAAADGKTYVIGN	1355
Dd	1300	TGARTINGQLLYFRANGVQVKGFEVTDYGRISYIDGNSGDIIRNRFVRNAQGMFVFDN	1359
Qy	1356	DGVAVDPSVVKQTFKDDASGALRFLYNLKGQLVTGSGWYETANHDWVYIQSGKALTGBOIT	1415
Dd	1360	NGYAV-----TGARTI	1370
Qy	1416	NGQHLFYKEDCHOVKGQVLTGTGDKVRYDANSQDAFNKSVTVNGKTYFYFGNDGTAQTA	1475
Dd	1371	NGQHLFYFRANGVQVKGFEVTDHGRISYIDGNSGDIIRNRFV-----	1412
Qy	1476	GNPKGQTFKDGDIRFYMSBEGQLVTGSGWYENAGQOMLYV--KNGKVLITGLQTVGSQRVYF	1534
Dd	1413	-----RNAQGMFVFDNNGYAVTGARTINGQHLYP	1442
Qy	1535	DENGIQAKKAVTSDGKIRYFDENSGSMI	1564
Dd	1443	RANGVQVKGFEVTDYGRISYIDANSGERV	1472
RESULT 8			
AAU98030			
ID	AAU98030	standard; protein; 1475 AA.	
XX	AAU98030;		
AC	AC		
XX	27-AUG-2002	(first entry)	
DT			
XX	S. mutans glucosyltransferase GTFB mutant I448V.		
DE			
XX	Glucosyltransferase; GTFB; transgenic plant; paper sizing;		
KW	coating composition; glucan; starch; latex; thermoplastic molecule;		
KW	amylolplast; vacuole; paper manufacture; mutant; mutein.		
XX			
OS	Streptococcus mutans.		
OS	Synthetic.		
XX			
XX	-Key	Location/Qualifiers	
FH	Misc-difference 448	/note= "Wild-type Ile substituted by Val"	
FT			
XX	US2002031826-A1.		
XX	14-MAR-2002.		
XX	19-DEC-2000; 2000US-00740274.		
XX	07-JUN-1995; 95US-00478704.		
PR	07-JUN-1995; 95US-00482711.		
PR	07-JUN-1995; 95US-00485243.		
PR	16-JAN-1998; 98US-00007999.		
PR	16-JAN-1998; 98US-00008172.		
PR	20-JAN-1998; 98US-00009620.		
PR	11-DEC-1998; 98US-00210361.		
XX	(NICH/) NICHOLS S. E.		
XX	Nichols SE;		
PI	WPI; 2002-414332/44.		
DR			
XX	Glucosyltransferase B or D protein useful for producing a glucan useful		
PT	as substitutes for and additions to modified starch and latexes in paper		
PT	manufacture, comprises mutations in specific positions.		
XX	Claim 36; Page; 44pp; English.		
XX	The invention an isolated protein comprising a glucosyltransferase (GTF)		
CC	B polypeptide having changes at position from I448V, D457N, D567T,		
CC	K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,		
CC	I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a		
CC	GTF D polypeptide having changes at positions from T589D, T589E, N471D,		

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 CC
 XX

SQ Sequence 1475 AA;

Query Match 54.8%; Score 4579; DB 5; Length 1475;
 Best Local Similarity 56.3%; Pred. No. 8.1e-269;
 Matches 895; Conservative 213; Mismatches 338; Indels 144; Gaps 18;

QY 1 MEKNVFKMHKKRWVTLVSASATMLASALGASVASADTDTA---SDSNQAVVTGQ 56
 DB 1 MDKKVYKURKKRWVTVSASAVMTLTLSSGLVKADSNESKQISDNTSVTANE 60
 QY 57 TTN--NQATDQTSIAATATSEQASTDAATQASAAEQGTGTASTDTAAQTITNANEAK 114
 DB 61 ESNVITEATSKQEAASSQTNHVTTSSTSVVNPK-----VVSNPVTGETASNGEKL- 115
 QY 115 WPTENENQGTDEMIAEA-----KNVATAESDSIPSD-----LAKSNVQVDGKY 161
 DB 116 -----QNTTVDKTSAAANNISQTTTEADTDVDDSNAAQLQLEKLPNYKEIDGKY 169
 QY 162 YYYDQGNVKNFAVSGDKIYYFDETGAVKDTSKVDADKSSASVQNTATPAANNRAYS 221
 DB 170 YYDDNNGKVRTNTLADGKILHFDGTGATDITSIDTVNK--DIVTTRSNLYKKYQVVD 227
 QY 222 TSAKNFEAVDNYLTADSWYRPKSLKDGKTWTESGKDDFRPLMAWMPDTETKRNVTNYM 281
 DB 228 RSAQSFEHVDHYLTASWYRPKYILKDGKTWTQSTEKDFRPLMTWMPDETQRQVNYM 287
 QY 282 NKVVGDITTYATSTQADLTAAAEIQAARLEQKITSNNTKWLRREALSAFVKTOPWNGE 341
 DB 288 NAQLGINKTYDDTSNQLQLNAAATIQAKTEAKITLTKNTDMLRQTISAFVKTSQAWNSD 347
 QY 342 SEKPYDHLQNGALLFDNQTDLTPDTQSNRYLLNRPTNCTGSLDSRFTYVNPNDPLGGYD 401
 DB 348 SEKPFDDHLQNGAVLDNBEKGLPYANSVRIINRPTNCTGKDPRTY--ADNTIGGYE 405
 QY 402 FLANDVDNSNPVQAEQLNWLHYLLNFGSIYANDADANPDSIRVDAVDNVDADLLQISS 461
 DB 406 FLANDVDNSNPVQAEQLNWLHFLMFGNIYANDPDANFDSVRVDAVDNVDADLLQIAG 465
 QY 462 DYLAAGYGDKNKNHNSVIVEAWSNDTPLYLHDDGNLMMNDKFRLSMLWSLAKPL 521

DB 466 DYLAAGYHKKNDKKAANDHLSILEAWSNDTPLYLHDDGNMNMNDKLSLLSFLAKPL 525
 QY 522 DKRSGNLPLIHNSLVDREVDREVTVPSSYPARAHDSVQDIIRDIIKAEINPNFSGYS 581
 DB 526 NQSGMNPILITNSLVNRITDDNAETAAPSPSYFIRAHDSVQDLIADIKAEINPNVVGYS 585
 QY 582 FTQBEIEQAFKLYNEDLKKTKYTHYNVPLSYTLITLTKGSIIPRYVYGDMDTDDGOYMA 641
 DB 586 FTMBEIKKAFETYNKDLATEKKYTHYNALSYALLTNKSSVPRVYVYGDMDTDDGOYMA 645
 QY 642 NKTVNYDAIESLKKARKMKVSGQAMONYQIGNGEILTSVRYGKALKGSKDGADATRTS 701
 DB 646 HKTINYEALETLLKARIKYVSGQAMRQOVNSEILTSVRYGKALKAKATDGTDRTRTS 705
 QY 702 GYGVVMGQNPNSLPGK--VVALNMGAAHANQBYRALMSTKDGVIATYATDADASKAGLVK 760
 DB 706 GVAVIEGNNPFLRLKASDRVVVMGAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
 QY 761 RTDENGILYFLNDDLKGVANPOVSGFLQVWVPVGAADDODIRVAASDTASTDGK--SLHQ 818
 DB 764 YTNDRGELIFTAADIKGVANPOVSGVLGVWVPVGA---LIKMFALRLARPHQOQMASVHQ 820
 QY 819 DAAMDSRVMEFGFSNFQSPATKEEYTNVVIANNVDKFSVSGITDFEMAPQVYSSTDGOF 878
 DB 821 NAALDSRVMEFGFSNFQAPATKKEEYTNVVIANNVDKFAENGWTFEMAPQVYSSTDGSP 880
 QY 879 LDSVIQNGYAFTRYDRLGMSKANKYGTADOLKAIKALHAKGLKVMADVPDQMTYFPKQ 938
 DB 881 LDSVIQNGYAFTRYDRLGMSKANKYGTADOLKAIKALHAKGLKVMADVPDQMTYFPKQ 940
 QY 939 EYVTVTRTDKPKPTAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKKEYPELTKKQ 998
 DB 941 EYVTVTRVDKYGTPVAGSQIKNTLYVVDGKSGKQQAQYGAFLQAKYPELFAKQ 1000
 QY 999 ISTGQAIDPSVKIKOWSAKYFNGSNILGRGADYVLSQDVSNKYFNVA--SDTLFLPSSLL 1056
 DB 1001 ISTGVPMDPSVKIKOWSAKYFNGTNILGRGAGYVLDQATNTYFNISDNKEINFLPKTLL 1060
 QY 1057 GKVESGTRYDGKGYIYNSSATGDQVKASFIITEAGNLYYFGKDGVMVTGAQTINGANYFF 1116
 DB 1061 NQDSQVGFSDGKGYVYIST--SGYQAKNTFISEGDKWYTFDNNGVMTVQAQSGINGVNYF 1119
 QY 1117 LENGTALRNTIYTDAGNSHYANDGKRYENGYYQOP--GNDWRYFKDGNMAVGLTTVDGNY 1175
 DB 1120 LSNGLQLRDAILKNEIDGTYYAYYNDGRRYENGYYQPMGVRHFNNGEMSVGLTVIDGQV 1179
 QY 1176 QYFDKDGVAQDKIIIVTRDGKRYRFDHNGNAATNTFIADKTGHWYLYLKGDKGVAVTGAOT 1235
 DB 1180 QYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNRFIENEESKWLILGEDGAATVGSQT 1239
 QY 1236 VGKQKLYFEANGQOVKGFVTSDEGKLYFYDVSDDMTDTFTFIEDKAGNWFYLGKGAAY 1295
 DB 1240 INGOLHYFRANGVQVKGEFVTDHGRISYDNGSGDQIRNRFRVNAQOGVFDNNGYAV 1299
 QY 1296 TGAQOTIRGQKLYFANGQOVKGDIVKGTGDKIRYDYDAKSGEQVFNTKTVRAADKTYVIGN 1355
 DB 1300 TGAQTINGOLLYFRANGVQVKGEFVTDYGRISYDNGSGDQIRNRFRVNAQOGVFDN 1359
 QY 1356 DGVAVDPSVVKGTQPKDASGALRFYNLKGQLVTGSGWYETANHWDVVIQSGKALTGBQTI 1415
 DB 1360 NGYAV-----TGARTI 1370
 QY 1416 NGOLHYFKEDGHQVKQLVGTGDKVRYDYDANSQDQAFNKSVTNKGTYFYFGNDGTAQTA 1475
 DB 1371 NGOLHYFRANGVQVKGEFVTDHGRISYDNGSGDQIRNRFRV----- 1412
 QY 1476 GNPKGOTFKDGSDIRFYSMEGQLVTGSGWYENACQWLYV--KNGKVLTLQTVGSQRYVP 1534
 DB 1413 -----RNAQOGVFDNNGYAVTGARTINGOHLIFY 1442
 QY 1535 DENGIOAKGKAVRTSDGKIRYFDENSGSMI 1564

Query Match		54.8%; Score 4575; DB 5; Length 1475;
Best local Similarity		56.3%; Pred. No. 1.4e-268;
Matches		895; Conservative 213; Mismatches 338; Indels 144; Gaps 18;
QY	1	MEKNRPFKHKVKKRWVTLVASATMLASALGASVASADTDTA-----SDDSNQAVTGDQ 56
DB	1	MDKKVRYLKKRWVTVSASAVMTLTLSGLVKADSNESKQISNDSNTSVWTANE 60
QY	57	TTN--NOATDQTSIAATPATSEQASDAADQASAAEQTGQTASTDTAAQTITTNANEAK 114
DB	61	ESNVITEATSKQBAASSQTNHTVTSSSTSVVNPKE-----VVSNPVTVGETASNGEKL- 115
QY	115	WVTENENQGTDEMLAEA-----KVVATAESDISPSD-----LAKMSNVKQVDGKY 161
DB	116	-----QNQTTTVDKTSAAAANNISKOTTEADTDVIDDSNAAILQILEKLPNKEIDGKY 169
QY	162	YYDQDQGNKFNFAVSVDGKIYFDETGAYKDTSKVDADKSSASVSNATIPAAINPAYS 221
DB	170	YYDNNKVRNTFTLIADGKILHFDEGTATDTSIDTVNK--DIVTTRSLYKXNQVYD 227
QY	222	TSANFEADVNYLTADSWYRPKSILKDGKTWTESGKDDFRPLLMAMPDPTETKRYNVYM 281
DB	228	RSASQSEHVDHLYLTAESWYRPKLKDGTWTQSTEKDPRPLMTWHPDQETQRYNVYM 287
QY	282	NKVVGIDKTYTASTQADLTAAELVQARIEQKITSNNTKWLREASIAFVKTPQWNCE 341
DB	288	NAQLGINKTYDTSNQLNLIAAATIQAKIEAKITLKNKDWLRQTISAFVKTSANSD 347
QY	342	SEKPYDHLQNGALLFDNQDLDLPDQSNVLLNRPTNQTGSLDSRFTYNNPNDPLGGYD 401
DB	348	SEKPFDDHLQNGAVLDNEGLTPYANSNRIILNRTPTNQTGKKDPRYT--ADNTTGGYE 405
QY	402	FLIANDVDSNPVVOAEQLNWLHYLNFSGSIYANDADANFDSIRVDADVNDVADLLQISS 461
DB	406	FLIANDVDSNPVVOAEQLNWLHFLNFGNIYANDPANFDSIRVDADVNDVADLLQIAG 465
QY	462	DYLKAAAGIDKNNKNNHVSIVEAWSNDPTPYLHDDGNLMMNDKPRLSMLWSLAKPL 521
DB	466	DYLKAAAGIHKNDKAANDHLSILEAWSNDPTPYLHDDGNMINDNKLRLSLFLSLAKPL 525
QY	522	DKRSLNPLIHLNSLDREVDDREVEVPSVSPARAHDSSEVODIIRLIIKAEINPNPFGVS 581
DB	526	NORSGMNPITLNSLNRDTONAETAAPVPSYFIRAHDSSEVODIADIIKAEINPNPVGVS 585
QY	582	FTOEIEQAFKIYNEDLKKTKKYTHYNVPLSYTLTLTNKGSIPRVYVYGDMDTDDQYMA 641
DB	586	FTWEEIKKAFIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGDMDTDDQYMA 645
QY	642	NKTVNYDAIESLLKARKMYVSGQAMQNYQIENGELITSVRYGKALKOSDKGDATRTS 701
DB	646	HKTINYEAETLLKARIKYVSGQAMRNQVGNSEIITSVRYGKALKATDGTGDRITRTS 705
QY	702	GVGVVNGQNPNSLDGK--VVALNMGAAHNOEYRALMVSTKQGVATYATDADASKAGLVK 760
DB	706	GVAVTEGNNPSRLKASDRVVMNGAAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
QY	761	RTDENGVLFLNDLKGANPOVSGFLQVWPVGAADDQDQIRVAASDTASTDGK--SLHQ 818
DB	764	YTNDRGELLFTAADIKGYANPQVSGYLGWVPVGA--LIKMFALRLARPHQOMASVHQ 820
QY	819	DAAMDSRVMEFGSNFSQFATKEEYTNVVIANNVDFKFSWGITDPEMAPQVYSSDQGF 878
DB	821	NAALDSRVMEFGSNFQAFATKEEYTNVVIANNVDFKFAEWGVTDFEMAPQVYSSDQGF 880
QY	879	LSVTONQVAFTRDYLGMKANKYGTADQLKAIKALHAKGLKNMADWVPQMYTFPKO 938
DB	881	LDSVIQNGVAFTRDYLGIKSNKYGTADDLKAIKALHSGKIKVMADWVPQMYTFPEK 940
QY	939	EVVTVTRTDKFKPTAGSINHSLYVTDTKSSGDDYQAKYGGAFDLDELKEYPELTKKO 998
DB	941	EVVTAIRVDKQITPVAGSQIKNTLYVVDGKSGKQQAQYKAGFALEELQAKYPELFAKQ 1000
QY	999	ISTGQAIIDPSVKIKQWSAKYFNGLGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1056
DB	1001	ISTGVPMDPSVKIKQWSAKYFNGLGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1060
QY	1057	GKVVESGIRVDGKGIYNSSATGDOVKASPIITEAGNLYYFKDGYMYTGAQTINGANYFP 1116
DB	1061	NQDSQVGSYDGGKGYIYST--SGYQAKNTFISEGDKWYFDDNNGYMYTGAQTSINGVNYF 1119
QY	1117	LENGTALNTIYTDACNHSYANDGKRYENGQQF--GNDWRYPKDCGNMAVGLTTVVGNV 1175
DB	1120	LSNGLQLRDAILKNEGTYAYYNGDGRYENGYYQFMSGVRHPPNNGEMSGLTVIDGOV 1179
QY	1176	QYFDKQGVQAKKIIIVTRDGKRYFDOHNGNAATNTFIADKTHWYLYLKGKGVAVTGAQT 1235
DB	1180	QYFDEMGYQAKGFVTTADGKIRYFDKSGNMYRNRFIENEZEGKWLVLGEDGAAVTGSQT 1239
QY	1236	VGQKLYFEANGQVKGDFVTSDEKLYFYDVDSGDMWTTDTTFIEDKAGNHYLKGKGAAN 1295
DB	1240	INGOHLVFRANGQVKGDFVTHHGRISYYDGSNGDQIRNFRVRNAQGMFYFDNNGYAV 1299
QY	1296	TGAQTIRGOKLYFKANGQVKGDIKGTGKIRYDAKSCEQVFNKTVKAADGKTYVIGN 1355
DB	1300	TGARTINGQLLYFRANGQVKGDFVTDRIYGRISYYDGSNGDQIRNFRVRNAQGMFYFDN 1359
QY	1356	DGVAVDPVSVVKGQTFKDGALRPFYNLKGQLVTGSGWYETANHDWYIYQSGKALTGBQTI 1415
DB	1360	NGYAV-----TGARTI 1370
QY	1416	NGOHLVFKEDGHQVKGQLVTGDKVRYVDANSQDQAFNKSVTVNGKTYFYFGNDGTAQTA 1475
DB	1371	NGOHLVFRANGQVKGDFVTDHGRISYYDGSNGDQIRNFRV----- 1412
QY	1476	GNPKGTQFKDGSIRFYSMEGQLVTGSGWYENAGQWLYV--KNGKVLTLQTVGSGORVYF 1534
DB	1413	-----RNAQGMFYFDNNGYAVTGARTINGOHLV 1442
QY	1535	DENGIQAKAVRTSDGKIRYFDENSGSMI 1564
DB	1443	RANGVQVKGDFVTDRIYGRISYYDANGSERV 1472
RESULT 11		
AAU98033		
ID	AAU98033	standard; protein; 1475 AA.
XX	AAU98033;	
AC	AAU98033;	
XX	27-AUG-2002	(first entry)
DT	27-AUG-2002	(first entry)
XX	S. mutans	glucosyltransferase GTFB mutant K1014T.
DE	S. mutans	glucosyltransferase GTFB mutant K1014T.
XX	Glucosyltransferase; GTFB; transgenic plant; paper sizing;	
KW	Glucosyltransferase; GTFB; transgenic plant; paper sizing;	
KW	coating composition; glucan; starch; latex; thermoplastic molecule;	
XX	amylolplast; vacuole; paper manufacture; mutant; mutein.	
OS	Streptococcus mutans.	
OS	Synthetic.	
XX	Key	Location/Qualifiers
PH	Key	Location/Qualifiers
FT	Misc-difference 1014	/note= "Wild-type Lys substituted by Thr"
XX	US2002031826-A1.	
XX	14-MAR-2002.	
XX	19-DEC-2000; 2000US-00740274.	
XX	07-JUN-1995; 95US-00478704.	
PR	07-JUN-1995; 95US-00482711.	
PR	16-JAN-1995; 95US-00485243.	
PR	16-JAN-1995; 98US-00007999.	
PR	16-JAN-1998; 98US-00008172.	
PR	20-JAN-1998; 98US-00009620.	

Qy	1296	TGAOTIRGOKLYFRANGOOVKGDITVKTGDKIRIYDAKSGBOVENKTKVKAADGKTYVIGN	1355
Db	1300	TGARTINGQLLYFRANGOVQVKGEPVTDYRGIRISYDGNSGDQIRNFRVRAQOGWFVFDN	1359
Qy	1356	DGVAIDPSVVKQTFKDGASGALRFPNLKGQLVTGSGWYETANHDMWYIQSGKALTGBQTI	1415
Db	1360	NGYAV----- : -----TGARTI	1370
Qy	1416	NGOHLXFKEDGHQVKGQLVTGTDGKRVYIDANSQDAFNKSVTVNGKTYFYGNDGTAQTA	1475
Db	1371	NGOHLXFRANGOVQVKGEPVTDHGRIRISYDGNSGDQIRNRPV-----	1412
Qy	1476	GNPKGQTFKQGSDIRFRFSMEQQLVTGSGWENAGOWLYV-KNCKVLTGLTGVGSQRVYF	1534
Db	1413	----- : -----RNaQGWFFYDNNNGYAVTGARTINGOHLXF	1442
Qy	1535	DENGIQAKGAVRTSDGKIRYFDENSGSMI	1564
Db	1443	RANGOVQVKGEPVTDYRGIRISYDANSGERV	1472

RESULT 12

AAU98032
ID AAU98032 standard; protein; 1475 AA.

AAU98032;

DT 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFB mutant D567T.

Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutin.

Streptococcus mutans.

OS Synthetic:

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

	Misc-difference	/note= "wild-type Asp substituted by Thr"
		567
FT		Location/Qualifiers

PN US2002031826-A1.

14-MAR-2002

19-DEC-2000: 2000US-00740274

XX 07-JUN-1995: 95HS-00478704

FK	07-JUN-1995;	95US-00478704.
PR	07-JUN-1995;	95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.
yy

PA (NICH/) NICHOLS S E.

XX
PI
Nichols SE:

WPI: 2002-414332/44.

Glucosyltransferase B or D protein useful for producing a glucan useful as substitutes for and additions to modified starch and latexes in paper manufacture, comprises mutations in specific positions.

PS Claim 36; Page; 44pp; English.

The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from I448V, D457N D567T.

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,

CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes p1 or p2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present properties represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

Query Match	54.8%	Score 4573;	DB 5;	Length 1475;
Best Local Similarity	56.3%;	Pred. No. 1.9e-268;		
Matches 895;	Conservative 212;	Mismatches 339;	Indels 144;	Gaps 18

```

Qy 1 MEKVRFPMHKVKRGWTTLSVASATMLASALGASVASADTDTA-----SDDSNOQAVVTGDQ 56
Db 1 MDKVRVYKLRVKCRWYTVSVASAVMTLTTLSSGLGVASASNESKSQISDNTSNTSVVTANE 60
Qy 57 TTN--NOATDOTSIAATATSEQASASTDAATDQASAAEQTGTTASTDTAAQTATTNANEAK 114
Db 61 ESNVITEATSKOEAASSQTHVTYTTSSSTSVVNPKE-----VVSNPYTVGSTASNGEKL- 115
Qy 115 WVPTEENOGQTDMLABA-----KNVATAESDISPSD-----LAKMSNVKQVGVKY 161
Db 116 -----QNQTTTVDVTKTSEAAANNISKQTTTEADTDOVIDDSNAANLQILEKLPNVKEIDGKY 169
Qy 162 YYYDODGNVKNFAVSGVDKIYFDEFGAYKOTSKVDADKSSSAVSQNAATIFAANNRAYS 221
Db 170 YYYDNNGKRVNFTLIADGKILHFDFTGATYDTSIDTVNK--DIVTTRSNNYKXKNQYVD 227
Qy 222 TSAKNFEAVDNYLTADSNYRPKSILKDGKWTWESGKODFRPLLMAWMPDTEKRNYNVM 281
Db 228 RSAQSPHEVDHYLTAESWYRPFVKILKDGKWTQSTEXKDFRPLMTWMPDQETQRQYNYVM 287
Qy 282 NKVVGIDKTYTATSQADLTAAAEVLVQAREOKTISENNYKWLBEATSAFVKTOPWNGE 341
Db 288 NAQLGINKTYDDTSNQLQNLAAATIQAKIEAKITLTKNTDMLRQTSFAFVKTOSAWNSD 347
Qy 342 SEKPYDHLQNGALLFDNQDLDTPDQTOSNYRLNRPNTNQTGSLDSRFTYVNPNDPLGGYD 401
Db 348 SEKPFDDHLQNGAVLYDNECKLTPYANSNYRILNRPNTNQTGKKDPRYT--ADNTIGGYE 405
Qy 402 FLIANDVDSNPVVOAEQLNWLHYLLNFGSIYANDADANFDSIRVDVNDVADLLQISS 461
Db 406 FLIANDVDSNPVVOAEQLNWLHFMFNIGNIYANDPANFDSIRVDVNDVADLLQIATG 465

```

QY 462 DYLKAAYGIDKNNKANNHVSIVEAWSNDNTPYLHDDGDNLMNMDNKPFLSLMLSLAKPL 521
DB 466 DYLKAAGKHNDKKAANDHLSILEAWSNDNTPYLHDDGDNLMNMDNKPFLSLMLSLAKPL 525
QY 522 DKRSLNPLTHSLVDREVDREVETVPYSFARAHDSVQDIIRDIIRKAEINPNPSFGYS 581
DB 526 NQRSGNPLTHSLVNRDNDNAETAAPVSPYSPIRAHDSVQDIIRDIIRKAEINPNVVGYS 585
QY 582 FTQEEIEQAFKLYNEDLKTDKKYTHYNVPLSYTLTLLTNKSGIPRYVYGMFTDDGQYMA 641
DB 586 FTWEEIKKAFIYNKDLATEKKYTHYNTALSYALLLTNKKSSVPRVYVYGMFTDDGQYMA 645
QY 642 NKTVNYDAIESLLKARMKVSGGQAMONYQINGEILTSRVYCKGALKQSDKGDAATTRS 701
DB 646 HKTINYEALETLLKARIKVSGGQAMRNOQVGNSEIITSRVYCKGALKATDGTDRTRTS 705
QY 702 GGVVVMGNPNFSLDGK--VVALNMGAHAHQEYRALMVSTKDGVAFYATDADASKAGLVK 760
DB 706 GVAVIEGNNPSELKASDRVVMGAAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
QY 761 RTDENGILYFLNDLKGVANPQVSGFLQVVPVGAADDQDIIRVAASDTASTDGK--SLHQ 818
DB 764 YTNDRGELIFTRAADIKGVANPQVSGVLGVWVPVGAA--LIKMFALRLARPHQQMASVHQ 820
QY 819 DAMDSRVNPEGSFNSQFATKEEYTNVVIANNVDKFSVSGITDPEMAPQVSSYSDGQF 878
DB 821 NAALDSRVNPEGSFNSQFATKEEYTNVVIANNVDKFAEWGVTDFEMAPQVSSYSDGSF 880
QY 879 LDSVIONGVAFTRDYLGMASKANKYGTADQLVKAIKALHAKGLKVMADVPQOMYTFPKQ 938
DB 881 LDSVIONGVAFTRDYLGLSGKPKYGTADDLVKAIKALHASKGKVMADVPQOMYAFPEK 940
QY 939 EYVTVTRTDKFKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKEKPELFTKKQ 998
DB 941 EVVTRVRYKGYTFVAGSQIKNTLYVVDGSSKQQAQYGGAFLEELQAKYPELFAKQ 1000
QY 999 ISTGQALDSVKIKQWAKYFNGSNILGRADYVLSDOVSNKYFNVA--SDTLFLPSSLL 1056
DB 1001 ISTGVPMDPSVKIKQWAKYFNGSNILGRAGYVLSKQATNTYFNISDNKEINFLPKTLL 1060
QY 1057 GKVVGSGIRYDGGYLYNSSATGDQVKASPIEAGNLXYFGKDGVNVTGAQTINGANYFF 1116
DB 1061 NQDSQVGFSDGKGYVYYS--SGYQAKNTFISEGDKWYFDNNGYVMTGAQSGINGVNYFF 1119
QY 1117 LENGTRALNTIYTDAGNSHYIYANDGKRYENGYQQF--GNDWRYFKDGNMAVGLTTVDGNV 1175
DB 1120 LSNGLQLRAILAKNEDGTVAYVNDGRRYENGYQQFMSGWVRHFNNGEMSVGLTVIDGQV 1179
QY 1176 QYFDDGVQAKDLIIVTRDGKVRYPFHNGNAAINTFIADKTHWYLYLKGDVAVVTGAQT 1235
DB 1180 QYFDENGYQAKGFVTTADGKIRYFDKQSGNMYRNFIEENEGKWLYLGEDGAAVTGSQT 1239
QY 1236 VGKQKLYFRANGQOVKGDVFTSDEGKLYFYDVSDDMTDTFTIEDKAGNWFYLGKDAV 1295
DB 1240 INCQHLYFRANGVQVKGFEVTDHGRISYDNGSGDQIRNFRVNAQGVFYFDNNGYAV 1299
QY 1296 TGAQTIRGQKLYFKANGQOVKGDIVKGTGDKIRYDASKSGEQVFNKTVKAADGKTYVIGN 1355
DB 1300 TGAQTIRGQKLYFRANGVQVKGFEVTDYGRISYDNGSGDQIRNFRVNAQGVFYFDN 1359
QY 1356 DGVAVDPSVVGQTFKDAASGALRFPYMLKGQLVTSWGYETANHDWYVYISGKALTGEOTI 1415
DB 1360 NGYAV-----TGARTI 1370
QY 1416 NGHLYFKEDGHQVKGQVLTGTDKVRYYDANSDDQAFNKSFTVNGKTYVYFGNDGTAQTA 1475
DB 1371 NGHLYFRANGVQVKGFEVTDHGRISYDNGSGDQIRNFRV----- 1412
QY 1476 GNPQGQTFKDGSDIRFYSMBGQLVTGSGWYENAGQWLXY--KNGKVLTLGLTVGSRVYF 1534
DB 1413 -----RNAQGVWYFDNNGYAVTGTATINGQHLXYF 1442
QY 1535 DENGIOAKGKAVTSDGKIRYFDNNGSGSMI 1564

DB 1443 RANGVQVKGFEVTDYGRISYDANSGERV 1472
RESULT 13
AAU98035
ID AAU98035 standard; protein; 1475 AA.
XX AAU98035;
XX 27-AUG-2002 (first entry)
XX S. mutans glucosyltransferase GTFB mutant D457N/D571K.
XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX Streptococcus mutans.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 457 /note= "Wild-type Asp substituted by Asn"
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
FT
XX US2002031826-A1.
XX 14-MAR-2002.
XX 19-DEC-2000; 2000US-00740274.
XX 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX Claim 36; Page: 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014I, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014I,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a

KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX Streptococcus mutans.
OS Synthetic.
XX Key Location/Qualifiers
FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"
FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
XX US2002031826-A1.
XX 14-MAR-2002.
XX 19-DEC-2000; 2000US-00740274.
XX 07-JUN-1995; 95US-00478704.
XX 07-JUN-1995; 95US-00482711.
XX 07-JUN-1995; 95US-00485243.
XX 16-JAN-1998; 98US-00007999.
XX 16-JAN-1998; 98US-00008172.
XX 20-JAN-1998; 98US-00009620.
XX 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX Claim 36; Page; 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper siring and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper siring/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer

CC	using the GTFB sequence appearing as AAU98027 and the information in
CC	claim 36
XX	
SQ	Sequence 1475 AA;
	Query Match 54.7%; Score 4568; DB 5; Length 1475;
	Best Local Similarity 56.2%; Pred. No. 3.8e-268;
	Matches 894; Conservative 213; Mismatches 339; Indels 144; Gaps 18;
QY	1 MEKNVRFKMHKVRWVTLVSASATMLASALGASVASADTDTA-----SDSNOAVVTGQ 56
DB	1 MDKVRKLRKVKRWRVTVSVASAVMTLTTLGGVLKADSNESKSI S NDSNTSVVTANE 60
QY	57 TTN--NOATDQTSIAATATSEQASASTDAATDOASAEQTGGTTASTDTAAQTITNANEAK 114
DB	61 ESNVITEATSKQEAASSQTNHTVTSSSTSVVNPKE-----VVSNNPYTGETASNGEKL- 115
QY	115 WYPTENENOGFTDEMLAEA-----KNVATAESDSIPSD-----LAKSNVQKVDGKY 161
DB	116 -----QNGTTTVDTSEAAANNISKQTEATDVIDDSNAANLQILEKLPNVEIDGKY 169
QY	162 YYYDQDGNVKNFPAVSVGDKIYYFDETGA YKDTSKVDADKSSASVQNTATIFAANNRAYS 221
DB	170 YYYDNNKVRTNFTLIADGKILHFDGTGAYTDTSDTVNK--DIVTTRSNLYKKNQVYD 227
QY	222 TSAKNFEAVDNYLTADSNWYRPKSILKDGKTWTESGKDDFRPILMAMWPDTEKCNVNYM 281
DB	228 RSAQSFEHVDHYLTAEBSWYRPKYILKDGKTWTQSTEKDPRLLMTWMPDQETQRQVNYM 287
QY	282 NKVVGIDKTYTAETSQADLTAAAEIVQARI EOKITSENNTKWLREAI SAPVKTPQOWNGE 341
DB	288 NAQLGINKTYDTSNQLQNLNIAAAIQAIEAKITLKTMDWLRQTISAFVKTSQAWSND 347
QY	342 SEKPYDDHLQNGALLFDNQDTLPTDPTQSNRYLNRTPNTQTSGLDSRFTYPNPDPLGGYD 401
DB	348 SEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRILNTPNTQTKGKDPRT--ADNTIGGYE 405
QY	402 FLLANDVNSNPVQAEQLNWLHYLLNFGSIYANDADANFDSIRVDADVNDADLQIIS 461
DB	406 FLLANDVNSNPVQAEQLNWLHFLNFGNIYANDPDANFDSIRVDADVNDADLQIAG 465
QY	462 DYLKAAYGIDKKNKXANNHVSIVEAWSNDNTPYLHDDGNLMNMDNKFSLMSLAKPL 521
DB	466 DYLKAAGLHKNDKKAANDHLSIDEAWSNDNTPYLHDDGNMNMNKLRLSLFLSLAKPL 525
QY	522 DKRSLNPLIHNSLVLDREYDREVETVPSPYSFARAHDSVQDIIRDIIKAEINPNPFGYS 581
DB	526 NQSGMNPILTNSLVNRTDDNAETAAPSPYSFIRAHDSVQTLIADI KAEINPNPVGYS 585
QY	582 FTQEEIEQAFKLYNEDLKKTDKKYTHYVPLSYTLILLTNKGSIPRVYVYGMFTDDGOYMA 641
DB	586 FTMEETKKAFAEYLNKDLATEKKYTHYNTALSYALLLTNKSVPFRVYVYGMFTDDGOYMA 645
QY	642 NKTVNYDAIESLLKARMKVSGQAMONYQIINGEILTSVRYGKAGKQSKDGDAITRTS 701
DB	646 HKTINYEALETLLKARIKTVSGQAMRNOQVNSEIITSVRYGKAGKALKATDGDTRTS 705
QY	702 GYGVVMGNQPNFSLDGK- YVALNMGAHAHQRYALMVSTKQGVATYATDADASKAGLVK 760
DB	706 GVAVIEGNNPSLRKASDRVVMGAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
QY	761 RTDENGYLFLNDLLKGVANPOVSGFLQVWPVGAADDODIRVAASDTASTDGK--SLHQ 818
DB	764 YTNDRGELIFTAADIKGANPOVSGYLVGMVVPVGA--L1KMFALURLARPHQOMASVHQ 820
QY	819 DAAMDSRVNFEFSGNFQSFATKEEYTNVIANNVDFVSWGITDFEMAPQYVSSYSDGQF 878
DB	821 NALDSRVNFEFSGNFQAFATKEEYTNVIAKNVDKFAEWGVTDEMAPQYVSSYSDGSP 880
QY	879 LDSVIQNGYAFTDRYDLGMSKANKYGTADOLVKAIKALHAKGLKKNMADWVPDQMYTFPKQ 938
DB	881 LDSVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKGIKNMADWVPDQMYAFPEK 940

QY 939 EVTVTRTKFGKPIAGSQINHSIYVTDTKSSGDDYQAKYGGAFDELKEKYPELFTKKQ 998
Db EVTVTRTKFGKPIAGSQINHSIYVTDTKSSGDDYQAKYGGAFDELKEKYPELFTKKQ 1000
QY 999 ISTGQATDPSVKIKOWSAKYPNGSNILRGADYVLSDOVSNKYENVA--SDTLPLPSLL 1056
Db ISTGVPMDPSVKIKOWSAKYPNGSNILRGADYVLSDOVSNKYENVA--SDTLPLPSLL 1060
QY 1057 GKVEGIRYDGKGIYINSSATFDQVKASFTITAGNLYIFCKDGYMTGATQINGANYFF 1116
Db NQDSQGVSGYDGRGVYVYST-SGVQAKNTFISEGDKWYFDNNGYMTGQASINGVNYYP 1119
QY 1117 LENGTALENTIYTDAGQNSHYANDGRYENGVOOF-GNDWRYFKQGNMAVGLTVTDGNY 1175
Db LSNGLQRLDAILKNEEDGTAYYNGDGRYENGYYQFMGVRHFNNGEMSVGLTVIDGQV 1179
QY 1176 QYFDKGVQAKKLIIVTRDCKVRVFDQHNAGNAATNTFIADKTGHVYLGKDGVAVTGAOT 1235
Db QYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNFIEENEGKWLVLGSDGAAVTGSQT 1239
QY 1236 VGRQKLYFEANGQOVKGFVTSDEGKLYFYDVSDDMTDTFTIEDKAGNWFYLGKDGAAV 1295
Db INGHLYFRANGVQVKGFEVTDHGRISYDNGSGDQIRNRFVRNACQGFYFDNNGYAV 1299
QY 1296 TGAQTIRGKLYFKANGQOVKGDIVKGTGDKIRYIDAKSGEQVFNKTVKAADGKTYVIGN 1355
Db TGAQTIRGKLYFRANGVQVKGFEVTDHGRISYDNGSGDQIRNRFVRNACQGFYFDN 1359
QY 1356 DGVAVDPVVKGTQFDKASALFYNLKGOLVTGSGHYETANDHWYIYQSKKALTGEOTI 1415
Db NGYAV-----TGARTI 1370
QY 1416 NGQHLFKEDGHQVKGOLVTGDKVRYDANGSDQAFNKSVTVNGKTYVFGNDGTAQTA 1475
Db NGQHLFRANGVQVKGFEVTDHGRISYDNGSGDQIRNRFV----- 1412
QY 1476 GNPKGOTFKDGSIRFYSMEGQLVTGSGWYENAGQWLYV-KNGKVLTLGLTVGSRQVYF 1534
Db -----RNAQGVWFYFDNNGYAVTGARTINGQHLV 1442
QY 1535 DENGIAQKAVRTSDGKIRYFDPENSGSMI 1564
Db RANGVQVKGFEVTDHGRISYDNGSGERV 1472

RESULT 15

AAU98036
ID AAU98036 standard; protein; 1475 AA.
AC AAU98036;
XX
XX
XX
XX 27-AUG-2002 (first entry)
DE S. mutans glucosyltransferase GTFB mutant D567T/D571K.
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; muten.
XX Streptococcus mutans.
OS Synthetic.

Key Location/Qualifiers
FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
XX US2002031826-A1.
XX 14-MAR-2002.
PD 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
PA Nichols SE;
XX WPI; 2002-414332/44.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
FT manufacture, comprises mutations in specific positions.
XX Claim 36; Page; 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX
XX Sequence 1475 AA;
SQ

Query Match 54.7%; Score 4566; DB 5; Length 1475;
Best Local Similarity 56.2%; Pred. No. 5e-269;
Matches 894; Conservative 212; Mismatches 340; Indels 144; Gaps 18;
QY 1 MEKNVRFKMHKVKCKWVTLVASATWLASALGASVASADTDTA----SDDSNOAVVTGDQ 56
Db 1 MDKKVRYKLRKVKCKWVTLVASAVMTLTLSGLVKADSNESKQISDNTSVVTANE 60
QY 57 TTN--NQATDQTSIAATATSEASATDAADQASAEQGTGTTASTDTTAQTTTANAEK 114
Db 61 ESNVITEATSKQEAASSQTNHTVTTSSTSSSTSVNPK-----VVSNNFTVGETASNGEKL- 115

```

115 WVPTENQFTDEMLAEA-----KXVATAESDSIPSD-----LAKMSNVKQVDGKY 161
116 -----QNOTTVDKTSSEAAANNISKOTTEADTDDVDDSSNAAMLQLEKLPNYKEIDGKY 169
162 YYVDQDNVKNFAVSVDGIYYIFDETGAYKOTSKYVDADKSSSAVSONATIPAAFNRAYS 221
170 YYVDNNGKVTNFTLADGKILHFDGTATDTSIDTVNK--DIVTTRSPLYKKNQOVD 227
222 TSAKNFEAVDNYLTADSWRPKSIKDGKTWTSEKDDFRPLMAMWPDTEKRNVTNYM 281
228 RSAQSPHEVDHYLTAESWRPKYILKDGKTWTQSTEKDFRPLMTWMPQDETQRQVVTM 287
282 NKVVGIDKTYTATSQADLTAAAEVLQAREIQKITSENNWTWMLREAISAFVKTOPOWNGE 341
288 NAQLGINKTYDDTSNQLNLNAAATIQAKIEAKITLLKNTDLWLRQTSIAFVKTSANSD 347
342 SERPYDDHLQNGALLFDNODTLDPTQSNYRLNRTPTNQTGLDSRFTVNPNDPLGGYD 401
348 SERPFDHLQNGAVLDNBEKLTIPYANSYRILNRTPTNQTGKDPRT--ADNTIGGYE 405
402 FLAANDVNSNPVQAEQNLNHLNFGSIYANDADANPDSIRVDAVDNVDADLLQIIS 461
406 FLAANDVNSNPVQAEQNLNHLNFGNIYANDPDANFDSIRVDAVDNVDADLLQIAG 465
462 DYLAAYGIDKNNKNNANHVSIIVEAWSDDNTPYLHDDGDMNMNDKFRLSMLWSLAKPL 521
466 DYLAAGAIGHKNDKAANDHLSILEAWSDDNTPYLHDDGDMNIMNDKRLSLFLSLAKPL 525
522 DKESGLNPLHNSLDREVDTVPVPSYSPARAHDSVQDIIIRDIKAEINPNPSFGYS 591
526 NQSGMNPILTNSLWRTDDNASTAAVPSYSFRAHDSVQTLIAKIIKAEINPNVVGYS 595
582 FTQEEIEQAFKIYNEDLKTKDKKYTHYNVPLSYLLTLTNKSGIPRVYVYGMFTDDGQYMA 641
586 FTWEEIKKAFIYNKDLATEKKYTHYNTALSYALLTLTNKSSVPRVYVYGMFTDDGQYMA 645
642 NKTVTNDAIESLLKARKYVSGGQAMQNYOIGNEILTSVRYGKGALQSKDGKDATRTS 701
646 HKTINYEAIETLLKARIKYVSGGQAMRNQOVGNSEIITSVRYGKGALKATDGTDRTRTS 705
702 GVGVMGNQPNFSLDGK-VVALNMGAAHANOEVYRALMVSTKDGCVATYATDADAKAGLVK 760
706 GVAIVIGNPESLKLKASDRVVMNGAAHKNQOATRPLLLTIDNGIKAYHSDQEA--AGLVR 763
761 RTDENGLYFLNDLKGVANPQVSGFLQVWVPVGAADDQDQIRVAASDTASTDGK--SLHQ 818
764 YTNDRGELIPTAADIKGYANPQVSGVLGVWVPVGA--LTKMFLRLARPHQOMASVHQ 820
819 DAAMDNRVMEGFSNFQSPATKEEYTNVVIANNVDKFSWGIITDFEMAPQYVSSTDGQF 878
821 NAALDSRVMEGFSNFQAFATKEEYTNVVIANNVDKFAEWGVTDFEMAPQYVSSTDGSF 880
879 LDSVIQNGYAFTDRYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMTTPKQ 938
881 LDSVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKVMADWVPDQMYAFPEK 940
939 EVVTVTRTRDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKEYPELFTKKQ 998
941 EVVTATRVDKYGTVPVAGSIKNTLYVVDGKSGKQQAQYGGAFLEELQAKYPELFARKQ 1000
999 ISTQOALDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1056
1001 ISTGVPMDPSPVKIKQWSAKYFNGTNILGRGAGYVLKDAQATNTYFNISDNKEINFLPKTL 1060
1057 GKVVESGIRYDGKGYLYNSSATGDQKASFIETAGNLYYFGKDGVMYVTAQTINGANYPF 1116
1061 NQDSQVGSYDGGGYVTYIST-SGQAKNTFISEGDKWYFDNNGYMWTAQSGINGVNYIF 1119
1117 LENGTALRNTIYTDAGNSHYANDGKRYENGYQQF-GNDWRYFKDGNMAVGLTTVDGNV 1175
1120 LSNGLQLRDAILKNEIDGTAYYGNDRRYENGYQFMGSGVWRHFNNGEMSVGLTVIDGQV 1179
```

```

1176 QYFDKQGVQAKDKIIIVTRDKVRYFDQHNMGNAATNTFIADKTGHWTYLLGKDGVAVTGAOT 1235
1180 QYFDEMGYQAKGKFTTADGKIRYFDKQSGNMYRNFIEENEGKWLJYLGEDGAANTGSQT 1239
1236 VGKQKLYFEANGQOVKGFVTSDEGLXFPYDVSDDMMTDTTIEDKAGNWFYLGKDGAAV 1295
1240 INGQHLYFRANGVQVKGEFVTDHGHGRISYYDGNSGDQIRNRFVRNAQGWFFYDNNGYAV 1299
1296 TGAQTIRGQKLYFKANGQOVKGDIVKGTGDKIRYDAKSGBQVFNKTVKAADKTVVIGN 1355
1300 TGARTINGOLLYFRANGVQVKGEFVTDYGRISYYDGNSGDQIRNRFVRNAQGWFFYFDN 1359
1356 DGVAVDPSVVKGTQFKDASGALRFYNLKGQLVTCGSGWYETANHDWVYIOSGRALTCGEQTI 1415
1360 NGTAV-----TGARTI 1370
1416 NGQHLIFYKBDGHQVKGQLVTGDKVRYYYDANDSGDQAFNKSVTVNGKTYFVFGNDGTAQTA 1475
1371 NGQHLIFYFRANGVQVKGEFVTDHGHGRISYYDGNSGDQIRNRFV-----1412
1476 GNPKGQTFKDGSDIRPYSMEGQLVTCGSGWYENAAQGWLVV-KNGKVLTLGLQTVGSGRVYF 1534
1413 -----RNAQGWFFYDNNGYAVTGARTINGQHLV 1442
1535 DENGIOAKGKAVRTSDGKIRYFEDENSGSMI 1564
1443 RANGVQVKGEFVTDYGRISYYDANDSGSERV 1472
```

Search completed: February 11, 2006, 19:14:59

Job time : 169.264 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 34.8119 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-37

Perfect score: 8349

Sequence: 1 MEKNVRFKMHKVKRWVTLS.....VYQYYFGNDGARIYRGMN 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/H_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4580	54.9	1475	2	US-09-007-999-2
2	4580	54.9	1475	2	US-09-210-361-2
3	4580	54.9	1475	2	US-09-740-274-2
4	4214.5	50.5	1375	2	US-09-210-361-4
5	4214.5	50.5	1375	2	US-09-740-274-4
6	3575	42.8	1430	2	US-09-008-172-2
7	3575	42.8	1430	2	US-09-210-361-6
8	3575	42.8	1430	2	US-09-740-274-6
9	3245	38.9	1577	1	US-08-793-824-2
10	2454	29.4	2057	2	US-09-499-203-2
11	2284	27.4	1278	2	US-09-604-957-3
12	2284	27.4	1781	2	US-09-995-749A-2
13	1653	19.8	545	2	US-09-604-957-4
14	1631	19.5	545	2	US-09-995-749A-10
15	1600	19.2	523	2	US-09-604-957-5
16	1595.5	19.1	522	2	US-09-604-957-11
17	1401	16.8	535	2	US-09-604-957-7
18	1401	16.8	535	2	US-09-995-749A-13
19	1260.5	15.1	584	2	US-09-995-749A-12
20	1255.5	15.0	584	2	US-09-604-957-6
21	694	8.3	349	2	US-09-009-620-2
22	514	6.2	2710	1	US-08-480-604A-6
23	514	6.2	2710	1	US-08-405-496A-6
24	514	6.2	2710	2	US-08-915-136-6
25	514	6.2	2710	2	US-08-957-310-6
26	514	6.2	2710	2	US-10-011-366-6
27	514	6.2	2710	2	US-09-084-517-6

28	459	5.5	866	2	US-09-545-773-2	Sequence 2, Appli
29	459	5.5	866	2	US-10-222-038-2	Sequence 2, Appli
30	450.5	5.4	811	1	US-08-480-604A-7	Sequence 7, Appli
31	450.5	5.4	811	1	US-08-405-496A-7	Sequence 7, Appli
32	450.5	5.4	811	2	US-08-915-136-7	Sequence 7, Appli
33	450.5	5.4	811	2	US-08-957-310-7	Sequence 7, Appli
34	450.5	5.4	811	2	US-10-011-366-7	Sequence 7, Appli
35	450.5	5.4	811	2	US-09-084-517-7	Sequence 7, Appli
36	450.5	5.4	812	1	US-08-480-604A-29	Sequence 29, Appli
37	450.5	5.4	812	2	US-08-915-136-29	Sequence 29, Appli
38	450.5	5.4	812	2	US-09-084-517-29	Sequence 29, Appli
39	388	4.6	2366	1	US-08-480-604A-10	Sequence 10, Appli
40	388	4.6	2366	1	US-08-405-496A-10	Sequence 10, Appli
41	388	4.6	2366	2	US-08-915-136-10	Sequence 10, Appli
42	388	4.6	2366	2	US-08-957-310-10	Sequence 10, Appli
43	388	4.6	2366	2	US-10-011-366-10	Sequence 10, Appli
44	388	4.6	2366	2	US-09-084-517-10	Sequence 10, Appli
45	351.5	4.2	1231	2	US-08-714-741-41	Sequence 41, Appli

ALIGNMENTS

RESULT 1

US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nicholas, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match 54.9%; Score 4580; DB 2; Length 1475;
Best Local Similarity 56.4%; Pred. No. 1.4e-313;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
QY 1 MEKNVRFKMHKVKRWVTLSVASATMLASALGASVASADTDTA-----SDDSNQAVVTGQ 56
Db 1 MDKVKYKLRKVKRWVTVSASAVMTLTLGGLVKADSNESKSIQINDSNTSVVTANE 60
QY 57 TTN--NQATDQTSIAATATSEQASDAATDQASAEQGTGTASTDTAAQTITNANEAK 114
Db 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKE-----VVSNPYTVGETASNGKL- 115
QY 115 WYPTENEGFTDEMLAE-----KNVATASDSIPSD-----LAKSNVKNQVDGY 161
Db 116 -----NQTTTVDKTSAAANNISKQTTEADTDVDDSNAAANLQILEKLPNKEIDGY 169
QY 162 YYYDQGNVKNFVSVGDKIYFDETAYKDTSKVDADKSSSAVSONATIPAAANNRAYS 221
Db 170 YYDNGKRVNTFTLIADGKILHFDETAYTDTSDTVNK--DIVTRNLNKKYNOVD 227
QY 222 TSAKNFEAVDNYLTADSWYRPSILKDGKTWTESGKDDFRPLLMAWPDTETKRNVTNYM 281
Db 228 RSAQSFHVHYLTAEWYRPSYKILKDGKTWTQSTQEKDFRPLMTWPPQETQRCVNTM 287
QY 282 NKVWIDKTYTAEQADLTAAAEVQARIQKITSNNKWLREAIASFVKTPQWNGE 341
Db 288 NAQLGINKTYDTSNQLQNLIAAATQAKIEAKITLKNTDWLRLQRTISAFVKTQSAWNS 347
QY 342 SEKPYDDHLQNGALLFDNQDTLTPDTQSNYRLNLTPTTQGTGSLDSRFTYNNPNDPLGGVD 401

Db 348 SEKFPDHLQNGAVLYDNEGKLTYPYANSNRIILNRTPTNQTGKKDPRYT--ADNTIGGYE 405
QY 402 FLIANDVDSNPNVQAEQLNWLHYLNFSGSIYANDADANFDSIRVDAVDNDADLLQISS 461
Db 406 FLIANDVDSNPNVQAEQLNWLHFLNFGNIYANDPDANFDSIRVDAVDNDADLLQIAG 465
QY 462 DYLLKAAIGIDKNKNANHHVSIVEASDNDTPYLHDDGDNLMMDKNFRLSMLSLAKPL 521
Db 466 DYLLKAAIGIHKNDKAANDHLSILEANSNDNTPYLHDDGDNMIMNDKNLRLSLFLSLAKPL 525
QY 522 DKRSGNLPLIHNLSLDREVDTVPYSYPARADHSEVODIIRDIKAEINPNPSFGYS 581
Db 526 NQSRSGNPLITNSLVNRDNDNAETAAVPYSYFIRAHDSVQDDIADIIRKAEINPNVVGYS 585
QY 582 FTQEEJEQAFKIYNEDLKTDKKYTHYNVPLSYTLTLLTNKSGISPRVYVYGMFTDDGQYMA 641
Db 586 FTMEEEKAFEIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVYGMFTDDGQYMA 645
QY 642 NKTVNDALIESLLKARKVYVGGQAMONYQIGNEILTSVRYGKGALKQSDKGDAATTRS 701
Db 646 HKTINYEATETLLKARIKYVGGQAMRNQVGNSEIITSVRYGKGALKATDGTGDRTRTS 705
QY 702 GUGVVMGNOPNPSLDGK-VVALNMGGAHANOBYRALMVSTKCGVATYATDADAKAGLVK 760
Db 706 GVAIEGNPNPSURLKASDRVVNMGAAHKNQAYRPLLLTTDNGIKAYHSDQEA--AGLVR 763
QY 761 RTDENGVLFLNDLKGXVANPQVSGFLQVWVPVGAADDQDQIRVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGYANPQVSGVLGVWVPVGA--LTKMFLARLARPHQOMASVHQ 820
QY 819 DAAMD SRVMEGSPNFSQSPATKEEYTNVVIANNVDFKVSWGITDFEMAPQVYVSSDQGF 878
Db 821 NAALDSRVNMEGSPNFSQSPATKEEYTNVVIANNVDFKFAEWGVTDFEMAPQVYVSSDQGF 880
QY 879 LDSVIQNGVAFTRYDYLGHSAKANKYCTADOLVKAIIKALHAKGLKVMADWVPDQMYFPKQ 938
Db 881 LDSVIQNGVAFTRYDYLGHSAKANKYCTADOLVKAIIKALHAKGLKVMADWVPDQMYAPPK 940
QY 939 EVTVTRTKDFKPIAGSQIHNLSLYVTDTKSSGDDYQAKYGGAFDLDELKPKYPELFTKKQ 998
Db 941 EVTVTRTKDYGTVPAGSQIKNTLYVVDGKSSGKQQAQYGGAFLEBQAKYPELFAKQ 1000
QY 999 ISTGQADPSVKTIQNSAKYFNQSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSLL 1056
Db 1001 ISTGVPMDPSVKTIQNSAKYFNQSNILGRGAGYVLDQAQNTYFNISDNKEINFLPKTL 1060
QY 1057 GKVESGIRVDGKGYLYNSSATGDQVKASPIEAGNLVYFGKQGYMVTGAQTINGANYFP 1116
Db 1061 NODSQVGFSDGKGYVYIST-SGYQAKNTFPISEGDKWIYFPDNNGYMVTGAQSIINGVNYFP 1119
QY 1117 LENGTALRNTIYTDAGNSHYIANDGKRYENGYQQF-GNDWRYFKDGNMAGVLTVDGNV 1175
Db 1120 LSNGLQLRDALXNEDGTAYIYNDGRRYENGYQFMSGVWRHFNNGEMSVGLTVIDGQV 1179
QY 1176 QYFDKQGVQAKDKIIVTRDGKRYFPOHNGNAATNTFIADKTHWHYLYGKDGVAVTGAQT 1235
Db 1180 QYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNFIEENEGKWLVLGSDGAATGSGQT 1239
QY 1236 VGHOKLYFEANGQVKGDFVTSDEKLYFYVDVDSGDMWTDTFEDKAGNWHFYLKGDAV 1295
Db 1240 INGQHLVFRANGQVKGDFVTHHGRISYYDNGSGGQIRNFRVNAQGGQYFYPDNGYAV 1299
QY 1296 TGAQTRGOKLYFKANGQVKGDIIVKGTGDKIRYDAKSQGEQFNKTVKAADGKTVYIGN 1355
Db 1300 TGARTINGQLLYFRANGQVKGDFVTDYGRISYIDGNSGQIRNFRVNAQGGQYFYPD 1359
QY 1356 DGVAVDPSVVKQTFKDDASGALRFPYNLKQLVGTSGWYETANHDWYVIOGSKALTEQTI 1415
Db 1360 NGYAV-----TGARTI 1370
QY 1416 NGQHLVFKEDGHQVKGQVLTGDKVRYXYDANSQDQAFNKSVTVNGKTYFPGNDGTAQTA 1475

Db 1371 NGQHLVFRANGQVKGDFVTDORHGRISYYDNGSGQDIRNRFV----- 1412
QY 1476 GNPKGQTFKQSGDIRFYSMEGOLVTGSGWYENAAQOWLYV-KNGKVLTLQIQTGVSQRYVF 1534
Db 1413 -----RNAQOWFYFDNNGYAVTGCARTINGQHLV 1442
QY 1535 DENGLOAKKAVRTSDGKIRYEDENSGSMI 1564
Db 1443 RANGQVKGDFVTDYGRISYYDANSGERV 1472
RESULT 2
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
Query Match 54.9%; Score 4580; DB 2; Length 1475;
Best Local Similarity 56.4%; Pred. No. 1.4e-313;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
QY 1 MEKNVRFKHKKVKKRWVTLVSASATMLASALGASVASADTDTA-----SDDSNQAVVTGDQ 56
Db 1 MDKVKRYKLRKKRWVTVSVASAVMTLTLSGGLVKADSNESKQISNDSNTSVVTANE 60
QY 57 TTN--NQATDQTSIAATATSEQASATDAATDAQAAEQGTGTTASTDTTAAQTNTTANAEAK 114
Db 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKE-----VVSNPYTVGETASNGEKL- 115
QY 115 WYPTENENQGTDEMLAEA-----KNVATAESDSIPSD-----LAKMSNVKQVODGKY 161
Db 116 -----QNTTIVDTKTESEAAANNISQTTTEADTDVDDSNAAMLQILEKLPNVEIDGKY 169
QY 162 YYYDQDGNVKKQNFVSGVGDKIYYFDETGYKDTISKYDADKSSSAVSQNAATIFAANNRAYS 221
Db 170 YYYDNGKVRTNFTLLADCKILHFDGTGAYTDTSIDTVNK--DIVTTRSNLYKKYQVYD 227
QY 222 TSAKNFEAVDNYLTADSWTRPKSILKDGKTTWETSGKDDFRPLLMAMPDPTTGTNTVNYM 281
Db 228 RSAQSPHEVDHYLTABSWRYPKYILKDGKTTQSTEKDFRPLMTWPDQETQORYVNYM 287
QY 282 NKVVGIDKTYTAETSOADLTAAABELVOAREIOKITSENKMLREALSAFVKTPQOWNGE 341
Db 288 NACLQINKTYDDTSNQLNIAAATQAKIEKITTLKNTDMLQRTISAFVKTQSAWNSD 347
QY 342 SKKPYDHLQNGALLFDNQTLTPTDQSNRYLNRFTNTQTSGLDSRFTPNPNDPLGGYD 401
Db 348 SEKFPDHLQNGAVLYDNEGKLTYPYANSNRIILNRTPTNQTGKKDPRYT--ADNTIGGYE 405

```
Qy 402 FLANDVDSNPVQABQLNLHLLNFGSIYADADANFDSIRVDVNDVADLLQISS 461
Db 406 FLANDVDSNPVQABQLNLHLLNFGSIYADADANFDSIRVDVNDVADLLQIAG 465
Qy 462 DYLKAAAGIDKNNKANNHVSIVKASNDNTPYLHDDGDMNNDKFRSLMLWSLAKPL 521
Db 466 DYLKAAAGIHKNDKAAADHLSILKASNDNTPYLHDDGDMNNDKFRSLMLWSLAKPL 525
Qy 522 DKESGLNPLIHNSLVDRVDEVTVPYSYFARAHDSVQDIIRDIKAEINPNSFGYS 581
Db 526 NQSGMNPPLTNSLVNRTDNDNAETAAPVSYSFIRAHDSVQDLIADIIKAEINPNVVGYS 585
Qy 582 FTQEEIEQAKIYNEDLKTKTKYTHYNVPLSYLLTNKGSIPRVYVGMFTDDGGYMA 641
Db 586 FTMEIEIKAFIEYNKDLLEKTKYTHYNVPLSYLLTNKGSIPRVYVGMFTDDGGYMA 645
Qy 642 NKTVNDYDAIESLAKARKYVSGGQAMQYQIENGELITSYRYGKALKQSKDGATRTS 701
Db 646 HKTINYEAETLLKARIKYVSGGQAMRQVGNSEIITSYRYGKALKATDGTDRTRTS 705
Qy 702 GUGVWGNQNFSLDGK- VVALNNGAAHANOYRALMVSTKDGATVATDADAKAGLVK 760
Db 706 GVAVIEGNNFSLBKASDRVVNNGAAHKNQATRPLLITDNGIKATHSQDEA--AGLVR 763
Qy 761 RTDENGYLEFLNDDLKGVANPOVSGFLQVWVPVGAADDQDIRVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGYANPOVSGYLGWVPVGA---LIKMFALRLARPHQOMASVHQ 820
Qy 819 DAAMDVRMPEGSNFSQSFATKBEETVNVVIANNDKFPVSWGITDFEMAPQYVSSDGGF 878
Db 821 NAALDSRVMEGFSNFQAFATKBEETVNVVIAKNNDKFAEWGVTDFEMAPQYVSSDGSF 880
Qy 879 LDSVIQNGVAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKMWADVPDQMYTFPKQ 938
Db 881 LDSVIQNGVAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKMWADVPDQMYTFPEK 940
Qy 939 EVVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFGLDELKEKYPELEFTKKQ 998
Db 941 EVVTVTRDKYGTVPAGSQIKNTLYVVDGKSSGKQQAQYGGAFLEELQAKYFELFARKQ 1000
Qy 999 ISTQALDPSVKIKQWASKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSLL 1056
Db 1001 ISTGVPMDSVKIKQWASKYFNGSNILGRGAGYVLSQVSNKYFNVA--SDTLFLPCKLL 1060
Qy 1057 GKVVESSIRVDGKGYLYNSSATGDQVKAPEITAGNLYYFGKQYVMTGAQTINGANYFF 1116
Db 1061 NODSQVGFSDGKGYLYYST-SGYQAKNTFISEGDKWYFPDNNGYMTGAQSGINGVNYFF 1119
Qy 1117 LENGTALRNTIYDAGNSHYANDGKRYENGYQQF-GNDWRVFKDGNMAVGLTVDGNV 1175
Db 1120 LSNGLQRLDALKNEDGTAYYNGDGRVYGYQPMGWRHFNNGESVGLTVIDGGV 1179
Qy 1176 QYPKDGVQAKKIIIVTRDGKRYFPOHNGNAATFIADKTHGWHYLYGKGVAVTGAQT 1235
Db 1180 QYFDENGYQAKGFVTTADGKIRYFKQSGNMYRNFIEENEGKWLVLGDEGAAVTGSQT 1239
Qy 1236 VGKQKLYFEANGQVKGDFVTSDEKLYFVDVDSGDWMTDTFTIEDKAGWYFLGKDGAAV 1295
Db 1240 INQGHLYFRANGQVKGDFVTDHGRISYVDGNSGQIIRNRFVRNAQGGWFYFDNNGYAV 1299
Qy 1296 TGAQTLRGKLYFKANGQVKGDI VKGTDGKIRYDAKSGEQVFNKVAADGKTVVIGN 1355
Db 1300 TGARTINGQLLFRANGQVKGDFVTRDGRISYVDGNSGQIIRNRFVRNAQGGWFYFDN 1359
Qy 1356 DGVAVDPSVVGQTFKPDASGALRFYNLKGQLVTGSGWYETANHDWYVYIQSGKALIGEORTI 1415
Db 1360 NGAV-----TGARTI 1370
Qy 1416 NGQHLFYKEDGHQVKGQLVTGTDGKRYVYDANSGDQAFNKSVTVNGKTYVFGNDGTAQTA 1475
Db 1371 NGQHLFYFRANGQVKGDFVTRDGRISYVDGNSGQIIRNRFV----- 1412
```

```
Qy 1476 GNPKGQTFKDGSDIRFYSMEGQLVTGSGWYENAGQGLYV-KNGKVLTLGLQTVGSQRVYF 1534
Db 1413 -----RNAQGGWFYFDNNGYAVTGARTINGQHLFY 1442
Qy 1535 DENGIOAKGAVRTSDGKIRYFDENSGSMI 1564
Db 1443 RANGVQVKGDFVTRDGRISYDANSGERV 1472

RESULT 3
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott B.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
```

Query Match 54.9%; Score 4580; DB 2; Length 1475;
Best Local Similarity 56.4%; Pred. No. 1.4e-313;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;

```
Qy 1 MEKVRFPQHKVKRWVTLVSASATMLASALGASASADTDTA-----SDDSNOAVVTGDQ 56
Db 1 MDKVRVYKLRVKRWVTVSVASAVMTLTLLSGGLVKADSNBSKQISNDSNTSVVTANE 60
Qy 57 TTN--NQATDQTSIAATATSEQSASDAATDQASAAEQGTGTASTDTAAQTNTANEAK 114
Db 61 ESNVTEATSKQEAASSQNTHTVTTSSSTSVVNPKE-----VVSNPYTVGETASNGEKL- 115
Qy 115 WYPTENENOGFTDEMLAEA-----KNVATAESDSIPSD-----LAKMSNVKQVDGKY 161
Db 116 -----QNQTTVTDKTSSEAAANNISKQTTTADTDVDDSNAAQLQILEKLPVKEIDGKY 169
Qy 162 YYDQDGNVKNFVSGDKIYYFDBTGAYKOTSKVDADKSSSAVSQNAITPAANNRAYS 221
Db 170 YYDNNNGKVRTFTLIADGKILHFDGTGAYTDTSIDTVNK--DIVTTRSNLYKKNQVYD 227
Qy 222 TSAKPEAVDNYLTADSWRYPKSIKDGKTWTESGKDDFERPLLMAMPDTETKRYNNYM 281
Db 228 RQAQSFHVDHLYTAESWYRPKYILKDGKTWTQSTEDKFRPLLMTPWPOETORQVNNY 287
Qy 282 NKVVGIDKTYTAAETSAQDLTAAALVQARIQKITSNNNTKMLREAIASFVKTPQWNGE 341
Db 288 NAQLGINKTYDDTSNQLQNLNIAAATQAKIEKITTLKNTDMLRQITAFVKTSQSNNSD 347
Qy 342 SKPKVDDHLLQNGALLFDNQTDLTPDTQSNRYLLNRTPTNQTGSLDSRFTYNNPNDPGYD 401
Db 348 SEKPFDDHLLQNGAVLYDNEGKLTVPVANSYRILNRTPTNQTGKDPRT--ADNTTIGYE 405
```

QY 402 FLLANDVDSNPVQAEQNLWHLNFGSIYANDADANFDSIRVDADVNDADLLOIS 461
Db 406 FLLANDVDSNPVQAEQNLWHLNFGNIYANDPANFDSIRVDADVNDADLLOIAG 465
QY 462 DYLAAGYGDKNKNANHHVSIYEAASDNDTPYLHDDGNLMMNKNFRLSMLSLAKPL 521
Db 466 DYLAAGYGHKNDKAANDHLSILEANSNDTPYLHDDGDMNMMNKNLRLSLFLSLAKPL 525
QY 522 DKRSGNLPLHNSLVREDDRETVPSYSFARAHSEVQDIIRDIKAEINPNFSGYS 581
Db 526 NQSRSGMPLITNSLVNRKTDNAETAAPVSYFIRAHDSVQDLIADIKAEINPNVVGYS 585
QY 582 FTQEBIEQAFKIYNEDLKTKDKKYYTHNVPLSYTHLLTNKSGIPRVYVYGMFTDDCOYMA 641
Db 586 FTWEEIKAFPEIYNKDLATEKYYTHNTALSVALLLTNKSSVPRVYVYGMFTDDCOYMA 645
QY 642 NKTWYDAIESLAKRMKYVSGQAMQNTQIGNEILTSVRYGKGALKOSDKGDAATTRS 701
Db 646 HKTINTEALTELKARIKYVSGQAMRNQVGNSEIITSVRYGKGALKATDGDTRTRTS 705
QY 702 GUGVMGNQPNFSLDGK - VVALNMGAAHANQETRALMVSTKDGVIATYATDADASKAGLVK 760
Db 706 GVAVIEGNNPFLKASDRVVNMGAAHKNQAYRPLLLTTDNGIKAYHSDQEA - AGLVR 763
QY 761 RTDENGILFLNDDLGKVPANQVSGFLOVWVPVGAADDQDIRVAASDTASTDGK - SLHQ 818
Db 764 YTNDRGELIPTAADIKGYANPQVSGYLGWVPVGA - LKMFALRLARPHQOMASVHQ 820
QY 819 DAAMDNRVMEFSGNFSQSPATKEEETNVVIANNVKFSWGITDPEMAPQVSSYTDGQF 878
Db 821 NAALDRVMEFSGNFSQAFATKEEETNVVIAKNVDKFAEWGVTDEMAPQVSSYTDGGSF 880
QY 879 LDSVIQNGYAFTDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFPKQ 938
Db 881 LDSVIQNGYAFTDRYDLGISKPNKYGTADDLVKAIKALHSGKIKVMADWVPDQMYAFPEK 940
QY 939 EVTVTRTKFGKPIAGSQINHSLYTDTKSSGDDYQAKYGGAFIDELKEKYPELTKKQ 998
Db 941 EVVTATRVKYGTFVAGSQIKNTLYVVDGKSGDKQAKYGGAFLEELQAKYKYPELFAKQ 1000
QY 999 ISGQAIQDPSVKIKQMSAKYFNGSNILGRGADVYVSDQVSNKYFNVA - SDTFLPLSSLL 1056
Db 1001 ISGVPMDPSVKIKQMSAKYFNGSNILGRGAGVYVLDQATNTYFNISDNKEINFLPKYLL 1060
QY 1057 GKVESGIRYDGGYIYNSSATGDQVKASFTPEAGNLYYFGKDGVMYVGTQTINGANYFF 1116
Db 1061 NQDSQVGFSDYKGVYVYST - SGVQAKNTFISEGDKWYFDDNNGYVMTGAQSGINGVNYF 1119
QY 1117 LENGTAIRNTIYDAQGNSHYYANDGKRYENGYQOP - GNDWRYFKDGNMAVGLTTVDGNV 1175
Db 1120 LSNGLQLRDAILKNEBDGTAYYNGDGRYENGYYQFMSGVYMRHFNNGEMSVGLTVIDGQV 1179
QY 1176 QYFDKQGVQAKDIIIVTRDGKRYFDOHNGNAATNTFIADKTGHVYVYLGKDGVAVTVGAOT 1235
Db 1180 QYFDEMGYQAKGFVTTADGKIIRYFKQSGMYMRNFIEENEGKNLYLGEDGAATVGSQT 1239
QY 1236 VGRQKLYFEANGQVQKGFVTSDEKLYFYDVSQDMWTDFTIEDKAGNWFYLGKDAV 1295
Db 1240 INGOHLYFRANGVQVKGFEVTDHGRISYYDNGSDQIRNRNFVRNAQGWYFDDNNGYAV 1299
QY 1296 TGAQTIRGKLYFKANGQVQKGDIVKGTDKIRYDYAKSGEQQVFNKTVKAADGKTYVIGN 1355
Db 1300 TGARTINGOLLYFRANGVQVKGFEVTDYGRISYYDNGSDQIRNRNFVRNAQGWYFDDN 1359
QY 1356 DGVAVDPVSVKQTFKXDGALRYNLKGLVGTSGWYETANHDWVYIQSGKALTGEQTI 1415
Db 1360 NGYAV - TGARTI 1370
QY 1416 NGQHLFYKEDGHQVKGQVLTGTDGKRYVYDANSGDQAFNKSVTYVNGKTYFNGDGTQTA 1475
Db 1371 NGQHLFYFRANGVQVKGFEVTDHGRISYYDNGSDQIRNRNFV - 1412
QY 1476 GNPKGQTFKDGSDIRFYFMSBEGQLVTGSGWYENAGQWLYV - KNGKVLTLGTLQVGSQRYVF 1534

Db 1413 -----RUAQGWFFDNGNGYAVTGARTINGOHLIF 1442
QY 1535 DENGIOAKGKAVRTSDGKIRYFDENSGSMI 1564
Db 1443 RANGVQVKGFEVTDYGRISYYDANSGERV 1472
RESULT 4
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match 50.5%; Score 4214.5; DB 2; Length 1375;
Best Local Similarity 60.3%; Pred. No. 7.4e-288;
Matches 811; Conservative 186; Mismatches 296; Indels 53; Gaps 14;
QY 1 MEKNVRFKHKKVKKRWVTLVSASATMLASALGASVADTDTASDSDNQAVVTDGT--T 58
Db 1 MEKKVRFKLRKKRWVTVSIASAVVTLTSLGSLVKAD---STDRQQAUTESQASLVT 57
QY 59 NQAVTDQTSIA---ATATSEQSASTDAATDQASAAEQTQGTAST--DTAAQTITNANEA- 113
Db 58 TSEAAKETLTATDTATSTATSQPTATVTDNVSTTNQSTNTTANTANFVVKPTTSEQAK 117
QY 114 -----KWVPTENEN---QGFTDEM-----LAEAKNVATAESDS 143
Db 118 TDNSDKIITTSKAVNRLTATGKFPVANNNTAHPKVTDKIVPIKPKIGKLQKPSLSQDD 177
QY 144 IPSDLAKMNVQVQDKYNYVDQGNVKNFAVSGDKLYYFDETCAYKDTSKVDADKSS 203
Db 178 IAA-LGNVNIIRKRVNGKYYKYKEDGTLQKNYALNNGKTFFFDETGALSNNT-LPSKKN 235
QY 204 SAVSQNATIPAAANNRAYSTSANKFNAVNVLTADSWYRPKSILKDGKWTESGKDDFRPL 263
Db 236 ITNNDNTNSFAQNVQYVSTDVANFEVDHYLTAESWYRKYILKDKGTWTQSTEKDFRPL 295
QY 264 LMAWPDPTKRNYNVNMKNVGVIDKTYTAETTSQADLTAAAEVLQARIKQITSENNTKW 323
Db 296 LMTWPDQSTQRQYVNMNAQLGIHQYNTATSTPLQLNLAAQTQTKIEKITAENKNTW 355
QY 324 LREAISAFVKTQDQWNGESEKPYDDHLQNGALLFDNQDTLTPDTQSNYRLNPTPTNQTG 383
Db 356 LRQTTISAFVKTQSAWNSDEKPFDDHLQGALLYSNNSKLTQSAANSYRLNPTPTNQTG 415
QY 384 SLDSRFTYNPDPGLGVDYFLLANDVDNSNPVQAEQLNWLHLLNFGSIYANDADANFDS 443


```
QY 684 GKALKQSKDGAATRTTSGVGVVGMGNQPNFSLDGK-VVALNMGAAHANQYERLALMVSTKD 742
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 714 GKALKATDGTGRTTSGVAVIEGNPNFLRKASDRVVVNMGAHKNOAYRPLLTIDN 773
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 743 GVAATADADAKAGLVKRTDENGILYFLNDDILKGVANPQVSGFLQVWPVGAADDDIR 802
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 774 GIKAYHSDQEA--AGLVRYNDRKELIFTAADIKGYANPQVSGILGVWPVGAADQDVR 831
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 803 VAASDTASTDGKSLHQAAMDNRVMEFGFSNFQSFATKEEBEYTNVVIANNVDFKFSWGIT 862
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 832 VAASTAPSTDGKSVHQNALDNRVMEFGFSNFQAFATKKEEYTNVVIANNVDFKFSWGIT 891
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 863 DEMAPQYVSSDTGQFLDSVINGYAFTRDYDLGMSKANKYGTADOLVKAIALHAKGLK 922
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 892 DEMAPQYVSSDTGSLDSVINGYAFTRDYDLGMSKANKYGTADOLVKAIALHAKGLK 951
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 923 VNADWVPDQMYTTPKQEVVTVTRTDKFGPKIAGSQIINHSLYVTDTKSSGDDYQAKYGGAP 982
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 952 VNADWVPDQMYALPEKEVTVTRTDKFGPKIAGSQIINHSLYVTDTKSSGDDYQAKYGGAP 1011
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 983 LDBLKEKYPELFTKQISTGQADIPSVKIKQMSAKYFNGSNILGRGADYVLSDOVSNKYF 1042
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1012 LEBLOAKYPELFPARKOISTGVPMDPSVKIKQMSAKYFNGSNILGRGADYVLSDOVSNKYF 1071
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1043 NVASDTFLPSSLLG-----KVVESGIRYDGKGIYVNSSATGDOVKASITEAGNLYYFG 1097
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1072 SLVSDNTFLPKSLVNPNGHTSSSVTGLVDFGKGYVYIST-SGNQAKNAFISLGNWYFD 1130
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1098 KQGYMTGAQTNGANYFLENGTALRNTIYTDACNSHYVYANDGKRYENGVOQFCNDWR 1157
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1131 NNGYMTGAQSIANGANYFELNSGIQRLNAYDNGNKVLSYVGNDRRYENGYLFCQQR 1190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1158 YFKDGNMAGLTVVGNVQYFVDKGVQAKDKIIVTRDGVRYPDQHNQNAATFTIADKT 1217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1191 YFQNGIMAVGLTRVHAGVYFDSAGFOAKQGFITTAGDKLRYFDRSGNQISNRFVRNSK 1250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1218 GHWYILGKGVAVTGAQTVGKQKLYPEANGQKVGDFVTSDEGLYFYVDVSDGMWTDTF 1277
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1251 GEWFLFDHNGVAVTGTVTNGQRLYFKPQGVQAKGSEFIRDANGYLRYDPSNGNEVRNRF 1310
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1278 IEDKACNWFYLGKGAAGVTCQAOTIRG 1303
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1311 VRNSKGEWFLFDHNGIAVTGARVUNG 1336
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; LATEXES IN PAPER MANUFACTURE
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match 42.8%; Score 3575; DB 2; Length 1430;
Best Local Similarity 48.4%; Pred. No. 9.1e-243;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

QY 1 MEKRVFQKHVKRVRVTLVSASA--TMLASALGASVAS-----ADTDITAS 44
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


Qy	1093	LYFGKCDGMYTGAQTINGANYFFLENGTALNTIYTDAQNSHHYANDCKRYENGVOQF	1152
Db	1127	WYFFDNNGHMYGLQQLNGEVOYFLSNGVQLRESFLENADGSKNYFGHLGNRYSNGYSF	1186
Qy	1153	GND--WRYF-KDGNMAVGLTTVDGNVOYFDKGVOAKDIIVTRDGVKRVYFDOHNGNAAT	1209
Db	1187	DNDSKWRYPDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITSGDGKKRVFDGSGNMVAV	1246
Qy	1210	NTFFADTKGHHYYLKGQAVTGAQTGVGKQLKYFEANGQQVKGFVTSDEGKLYFYDVDS	1269
Db	1247	NRFANDKNGDWYYLNSDGIALVGQTINGKTYYYFGQDGKQIKGKIIT-DNGKLKYFLANS	1305
Qy	1270	GDMWTDTEFIEDKAGNWFVLGKDGAAVTCAGTIIRGKLYFKANGQOVKGDIVKCTDGKIRY	1329
Db	1306	GELARNIFATDSQNNWYFFGSDGVAVTGSQTIAGKULYFASDGQVKGVSFVT-YNGKVHY	1364
Qy	1330	YDAKSGEQVFNKTKAADGKTYIVTGNBGVAVDPSVVKGQTFKQASGALRFYNLKGQIVTG	1389
Db	1365	YHADSJELQVNRFEADKDG-----	1383
Qy	1390	SGWYETANHDIWVYIQS-GKALTGQTINGQHLYFKBDGHQVKGQL	1433
Db	1384	-----NWYILDSNGEALTQSQRINDQRVFTREGKQVKGDV	1419

```

RESULT 7
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
; US-09-210-361-6

```

```
Query Match      | 42.8%; Score 3575; DB 2; Length 1430;
Best Local Similarity 48.4%; Pred. No. 9.1e-243;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;
```

Qy	1	MEKNVRFGKHVKKRWVTLVSVA--TWLASALCASVAS-----ADDTAS 44
		: : : : : :
Db	1	METKRYIKMHKKHWHVVAVASGLITLGTLLGSSVSAETEQTSDKVTKSEDDKAA 60
		: : : : : :
Qy	45	DDSQA---VVTGDQTTNNOATDTSIAATATS-EQSASTDAATDOASAABQTGGTAST 100
		: : : : : : : : : :
Db	61	SESSQTDAPKTQAQTEQTAQSQANVADTSITKETPSQNIIITQANSDDKTVNTKSE 120
		: : : : : : : : : :
Qy	101	D-TAAQTNTTNANEAKVPTENENQGFTDEMLAEAKNVAT-AESDSIPSLAKMNVNQV 157
		: : : : : : : : : :
Db	121	EAOFSERTKQSEBAQ---TTASSCALTPKAELITKORQTAOANKNPVDLAALPNVKOI 177
		: : : : : : : : : :

[illegible]

Db 1247 NRPANDKNGDWYLYNSDGLALGVQTINGKTYFFQDGGKQIKGKIIT-DNGKLYFLANS 1305
QY 1270 GDMWTDFTIEDKAGNFWYLGKDGAAVTAQTIRGQKLYFKANGQOVQKGVIVKGTGDKIRY 1329
Db 1306 GELARNIFATDSQNNWYFGSDGVAVTGSQTAGKLYFASDGKQVKGFSVFT-YNGKVHY 1364
QY 1330 YDAKSGEQVFNKTVKAADGKTYVYIGNDGVAVDPSPVVGKQTFKDGASGALRFPYNLKGOLVTG 1389
Db 1365 YHADSGLQVNRFEADKDG-----1383
QY 1390 SGWYETANHDWYIOS-KHALTCEQTINGOHLFYKEDGHQVKOL 1433
Db 1384 -----NWYLDNSGEALTGSQRINDQRVFFTRGKQVKGDV 1419

RESULT 8

US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match 42.8%; Score 3575; DB 2; Length 1430;
Best Local Similarity 48.4%; Pred. No. 9.1e-243;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Caps 26;
QY 1 MEKNVFKKHVKVKKRWVTLVSASA--TMLASALGASVAS-----ADTDTAS 44
Db 1 METKRYKKHKKHKKWVTVAVASGLITLGTTLIGSSVASAEOTEQTSKVKVTKSEDKAA 60
QY 45 DDSNQA---VVTGDTNNQATDQTSIAATATS-EQASASTDAATDQASAAEQQTGTTAST 100
Db 61 SESSQTDAPKTKQAQTEQTAQSQANVADTSTSIKTEPSONITQANSDDKTVTNKSE 120
QY 101 D--TAQOTTNANEAKWPTENENQGTDEMLAEAKNVAT-AESDSIPSDLAKMSNVKQV 157
Db 121 EAQTSBERTKQSEAQ---TTASSQALTOAKELTKQRTAAQENKNPVDLAAIPNVKQI 177
QY 158 DGKYYVDDGDNVKKNFVSGDKIYYFDE-TCAYKDTSKVDADKSSVASQONATIPAA 216
Db 178 DGKYYIGSDGQPKKFNALVNNKVLVFDKNTGALTDTISQYQKGLTKLND---YTH 234
QY 217 NRAYSTSAKNFEAVDNYLTADSWYRPRKSLKDGKWTWESGKODFRPLLMAWPDTETGRN 276
Db 235 NQIVNFENTSLETIDNVVTADSWYRPKDILKNGKWTWASSESDLRPLLMASWPKDQTOIA 294
QY 277 YVNMKV-VGIDKTYTAETSQADLTAAAEVLQARIEQKITSNTNKLREALSAFVKIQ 335

Db 295 YLYNWNQOGLGTGENYTDASSQESLNLAQVQVQKIETKISQTOQOWLRDINSFVKQ 354
QY 336 POWNGESEKPYD---DHLQNGALLFDNQTDLTPTQSNRYLLNRTPTNQTSLSRFTY 391
Db 355 PNWNSQTESDTSAGEKQHLQGGALLYSN-SDKTAVANSDYRLNRTPTSQG---KPKY 409
QY 392 NPNDPLGGYDFLLANDVNSNPVQABQLNWLHYLLNFGSIYANDADANFDSIRVDADV 451
Db 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNWLHYLMNYGSIYANDPEANFDFGVRVDADV 469
QY 452 VDADILQJSSDYLKAAYGIDKNNKNANHVSIVAEASDNDTPYLHDDGDNLNMDNKPFL 511
Db 470 VNADLLQJASDYLKAHYGVDSKNAIHLISILEASDNDPOYNKDTKGAQAPIDNKLRL 529
QY 512 SMLSLAKPLDK-----RSLNPLIHNSIADREVDDRETVETVPYSFARAHSEVQD 563
Db 530 SLLYALTRPLEKQASNKNEIRSGLEFPVITNSLNNSAEGKNSERMANYIFIRAHSEVQT 589
QY 564 IIRDIKAEINPNSFGYFTQBEIEQAFKIYNEDLKKTKKTHYNNPLSYLLLTNKG 623
Db 590 VIAKIIKAQINPKTDGLTFLDELQAKFIYNEDMROAKKKTQSNIPTAYALMLSNKDS 649
QY 624 IPRVYVYGMFTDDGOYMAKNTVNYDAIESLLKARKKYVSGGOAMO-NYQIENG 676
Db 650 ITRLYIGMYSDDGQYMATKSPYDAIDTLLKARIKYAAGQDMKITYYEGDKSHMDWDY 709
QY 677 --ILTSVRYGKALKQSDKGDATRTSGVGVGMQNPFLD-GKVVALNMGAHAHQY 733
Db 710 TGVLTSVRYGTGANEATDQSEATKTQGMVAVITSNPNSLKNQNDKVIYVMGAHKNQY 769
QY 734 RALMVSTKDGVAITYATDADAKAGLVKRTDENGILYFLNDDLKGVANPOVSFGLQVWVPV 793
Db 770 RPLLLTTKDLGTSYTSDAAK--SLYRKTNDKGLVDFDASDIQGYLNPQVSGYLAWVVPV 827
QY 794 GAADDQJRVASDASTASTDQKSLHODAAAMDSRVMEFGFSNFQSEATKEEYTNVVIANNV 853
Db 828 GASDNDQVRVAASNKANATQVYESSALDSQLITGEGFSNFQDFTVKDSYTNKKIAQNV 887
QY 854 DKFVSWGITDFEMAPQYVSSSTDQFLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKA 913
Db 888 QLFKSWGVTSEFMAQYVSSSEDSFLDSIIQNGYAFEDRYDLAMSNNKNGYSGQDMINAV 947
QY 914 KALHAKGLKVMADWVPDQMYTPPKQEVTVTTRDFKFGFIAGSQINHSILYVTDTKSGDD 973
Db 948 KALHKSQIQVIADWVPDQIYNLPKQEVTVATRVNDYGEYRKDSEIKNTLYAANTKSGKD 1007
QY 974 YQAKYGGAPLDELKEKYPELFTKKOISTGQAITDPSVKIKQWSAKYFNGSNILGRGADYVL 1033
Db 1008 YQAKYGGAPLSELAARYPSIFNRTQISNGKIDPSEKITAWAKAYFNGNILGRGVYVL 1067
QY 1034 SDQVSNKYFNVASDTLFLPSSLLGKVVESGIRYDGKGYIYNGSSATGDQVKASFIIEA-GN 1092
Db 1068 KDNASDKYFELKNGNTYLPKQMTNKEASTGFVNDGNGMTFYST-SCYQAKNSFPVQAKGN 1126
QY 1093 LYVFGKDGWMTGAOTINGANFYFLENGTALNWTIYTDAGNSHYIYANDKRYENGYYQOF 1152
Db 1127 WTYFNNHGMVYGLQOLNGEVQYFSLNGVQLRESFLENADGSKNYFCHLGNRYSGYYSF 1186
QY 1153 GND--WRYF-KDGNMVAVLTTVDGNVQYFDKGVQAKDKIIVTRDCKVRYFDQHNGNAAT 1209
Db 1187 DNDKRWYFADSGVMVAGLUKTNGNTQYFDQDQYQVKGAWITGSDGKRYFDDGSGNMAV 1246
QY 1210 NTFIADKTHWYTLGKDGVAVTGAQTVGKQKLYFEANGQOVKQDGFVTSDEGLYFYDVDS 1269
Db 1247 NRPANDKNGDWYLYNSDGLALGVQTINGKTYFFQDGGKQIKGKIIT-DNGKLYFLANS 1305
QY 1270 GDMWTDFTIEDKAGNFWYLGKDGAAVTAQTIRGQKLYFKANGQOVQKGVIVKGTGDKIRY 1329
Db 1306 GELARNIFATDSQNNWYFGSDGVAVTGSQTAGKLYFASDGKQVKGFSVFT-YNGKVHY 1364
QY 1330 YDAKSGEQVFNKTVKAADGKTYVYIGNDGVAVDPSPVVGKQTFKDGASGALRFPYNLKGOLVTG 1389

Db 1365 YHADS6ELQVNRFRADKDG-----1383

Qy 1390 SGMETANHDVYIQS-GKALTGEQTINGQHLYFKEDGHQVKGOL 1433

Db 1384 -----NWYLDNNGEALTGSORINDQRVFFTRREGKQVKGDV 1419

RESULT 9

US-08-793-824-2

; Sequence 2, Application US/08793824

; Patent No. 5981838

; GENERAL INFORMATION:

; APPLICANT: Simpson, Christine Lynn

; APPLICANT: Giffard, Philip Morrison

; APPLICANT: Jacques, Nicholas Anthony

; TITLE OF INVENTION: Genetic Manipulation of Plants to

; TITLE OF INVENTION: Increase Stored Carbohydrates

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Griffith Hack & Co

; STREET: Level 8, 168 Walker Street

; CITY: No. 5981838th Sydney

; STATE: New South Wales

; COUNTRY: Australia

; ZIP: 2060

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/793,824

; FILING DATE:

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: AU PM7643

; FILING DATE: 24-AUG-1994

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 61 2 9957 5944

; TELEFAX: 61 2 957 6288

; TELEX: 26547

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1577 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: not relevant

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Streptococcus salivarius

US-08-793-824-2

Query Match 38.9%; Score 3245; DB 1; Length 1577;

Best Local Similarity 42.0%; Pred. No. 1.9e-219;

Matches 714; Conservative 236; Mismatches 486; Indels 264; Gaps 34;

Qy 1 MENKVRFKHKKVKKRWVTLVSASATMLASALGASVA-----SADTDTA-----SDDS 47

Db 1 MENKVRFKHKKVKKRWVTLVSALAGSLAQKQVEADETSAPNGDGLQQLSEDG 60

Qy 48 NQAVVTGDDQTNQATDQTSIAATATSEQS-----ASTDAATDQASAAEQTGTTASTD 101

Db 61 TASLVT-TTIVTEQASQASVSASVATSVSHETSFOAATSASQEATAQAT-SPVASQE 118

Qy 102 TAAQTNTANEAKWPTENENQGFDEMLA-----BAKNVATA 139

Db 119 VAVSSQTSSGQETOTTEQVSQGTSTQVAGQTSQAQSTPSVTEQARPRVLTNAAPAIATR 178

Qy 140 ESDSI-----144

Db 179 AADSTRINARNRNTNITASGTPNVTIITGNTPKPNVTVTSPNGTRPNVTIVTPNQ 238

Qy 145 -----PSDLAK-----MSNVKQVDGKYYYYDQDGNVKKNFASVSGDK 181

Db 239 PNKEVOPSPQSPQNPQVPQNPQSLDYKPVASNLKLTIDGKQYV-ENGWVVKNAIELDGR 297

Qy 182 IYFDETGAYKDTSK-----VDADKSSSAVSONATIPANNRAYSTSAKNFEAVDNYLTAD 237

Db 298 LYFDETGAMVDQSKPLRAD-----IPNNSIYAVYQAQYDTSKSEHLNDFLTAD 350

Qy 238 SWYRPSILKDGKWTWESGKDDFRPLLMAMWPDTEKRNNTYNNYNNKVVGVGIDKTYTAETSQ 297

Db 351 SWYRPSILKDGKWTWESGKDDFRPLLMAMWPDTEKRNNTYNNYNNKVVGVGIDKTYTAETSQ 410

Qy 298 ADLTAALVQVARIQKITSNNYNNKWLREAIASAFVKTQPOWNGESEKPY---DDHLONGA 354

Db 411 YDLAAAAETVQGIIEEREGNTTWLRQLMSDFIKTPQGNWSESDNLLVKGDLHOGGA 470

Qy 355 LLDPNQDLPDTPQSNRYLNRPTNCTGSLDSRFTYNNPDPLGGYDFLLANDVNSNPV 414

Db 471 LTFNNNS-ATSHANSDFRLNRPPTNCTGTR-----KYHIDRSNGGYELLANDIDNSNPA 525

Qy 415 VQAEQLNWLHYLLNFGSIYANDADANFDSIRVDADVNDADLLQIISDYLLKAAAGIDKN 474

Db 526 VQAEQLNWLHYLLNFGSIYANDADANFDSIRVDADVNDADLLQIISDYLLKAAAGIDKN 585

Qy 475 KNANHVSIIVASDNDPTPYLHDDGDNLMNMKNFRLSMLWSLAKPLDKKSGSLNPLTHNS 534

Db 586 ANAIAHLSTLEAWSYNDHQYNKDTGAQLSIDNPLRETLTTLFLRKSNNYRGSLSERVITNS 645

Qy 535 LVDRVDDREVEVTPVPSYFARAHDSVQDIIIRDIKAEINPNSFGVSTOBEETQAPKIY 594

Db 646 LNNRSEQRHTPRDANYIFVRHDSVQAVLANIISKQINPKTDGFTTMDLQQAPEIY 705

Qy 595 NEDLKEKTKKTYHNVPLSYTLTLTNKGSIPRVYVYGMFTDDQGYMANKTVNDAISLL 654

Db 706 NADIAKADKKYQYNIIPAAVATWLTNKSITRVYVYGDLPFTDDQGYMAEKSPYNAIDALL 765

Qy 655 KARKVTVSGQAMQNYQINGEILTSVRYKGGALKQSGDKGADATRTSGVGVVMGNQNF 714

Db 766 RARIKYVAGQDMKVTKLNGYEIMSSVRYKGAEEANQLGTAETRNQGLVLTANRRPDMK 825

Qy 715 LDGK-VVALNMGAAHQAQVRAVMSTKQGVATYATDADAKAGLVKRTDENGVIYFLND 773

Db 826 LGANDRLVNMGAHKNQAYRPLLLSKSTGLATYLDKSDV-PAGLVRYTDNQGLNLTAD 884

Qy 774 DLKGVANPQVSGFLQVWVPVGAADDDQIRVAASDTASTDCKSLHQDAAMDSRVMEFGSN 833

Db 885 DIAGHSTVEVSGYLAWVPVGAASENQDARTKASTKKE-QVFESSAALDSQVIYEGFSN 943

Qy 834 FQSFAKKEBYTNVVIANNVDFVSVGIDTFEMAPQYVSTGDFLDSVQIONGYAFVDRY 893

Db 944 FQDFVKTPSQYTNRVIAQNAKLFKEWGITSFEPAPQYVSSQDGTFLDSIENGVAPEDRY 1003

Qy 894 DLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPMQYTFPKQEVVTVTRTDKFKPI 953

Db 1004 DIAMSKNNKYSLKLDLMDALRALHAEGISAIDAMPDQIYNLPKQEVVTVTRTDKFKPI 1063

Qy 954 AGSQIHSLVYVDTKSGDDYQAKYGAFLDELEKYPFLTKKQISTGQAIQIPSVKIQ 1013

Db 1064 PHAEIYNSLYAKTRTFGNDPQKYGGAFLDELKAKYPAIFERVQISNGRKLTTNEKITQ 1123

Qy 1014 WSACYFNGSNI LRGADYVLSQVSNKYFNVAASDTLFLPSSLLGKGVESGIRYDGKGIY 1073

Db 1124 WSACYFNGSNI LRGADYVLSQVSNKYFNVAASDTLFLPSSLLGKGVESGIRYDGKGIY 1182

Qy 1074 NSSATGDDQKASPIEAGN-LYFEGDKGVMVTCAQTINGANYFLENGTALRNTIYTDQ 1132

Db 1183 -LSIGGYLAKNTPIQVGAQWYFDKXGNNVTGEQVIGDKKFFLQGLRHLVLRQGS 1241

Qy 1133 GNSHYVANDGKRYEYQOFG---NDWRYFKDGN---MAVGLTVDGNVQYFD-KDGVOAK 1186

Db 1242 GHVYVYDQKGVAFNGFYDFAGPRQDVRIF-DGNGQMYRGLHDMYGTTFDFDEKTIQAK 1300

Qy 1187 DKIIIVTRDGKRYFDQHNNGNAATNTFIAD-KTGHVYLLGKGVAVTGAQTVGKQKLYFEA 1245

Db 1301 DKFIRPADGTRFYFIPDTGNLAVNRFAQNPNKAWYYLDSNGYAVTGLQTQKQYYFDN 1360
Qy 1246 NGQOVKGD FVTSDEKLYFYVDVSGDMWTDTFIEDKAGNWFYLGKDGAAVTAQTIRGQK 1305
Db 1361 EGRQVKGHFVTINNQR-YFLDGSGBTAPSRFVTEN-NKWWYYVDGNGKLVKGAQVINGNH 1418
Qy 1306 LYFKANGQOVKGDIVKGTGDKIRBYDAKSGEQVFNKTVKRAADGKTYVIGNDGVAVDPSVV 1365
Db 1419 YFFNNDYSQVKGAWANG-----RYDGDGSGQAVSNQFI----- 1451
Qy 1366 KGQTFKADSGALRFYNLKGQVLTGSGWYETANHDWYI-OSGKALTEGQTINGQHLYFKE 1424
Db 1452 -----QIAANQWALNQDGHKVTGLQNINNKVYFEGS 1483
Qy 1425 DGHQVKGQVLTGDKVRYVDANSQDAFNKSVTVNGKTYFYFNGDGTAGTAPGNKQGTFFK 1484
Db 1484 NGAQVKGKLLT-VQGKKCYFDAHTGEQVNRV----- 1515
Qy 1485 DGSDFRYSMEGQVLTGSGHYENAGQWLYVKN-GKVLTLGLOTVGSQRYVFDENGIOAKG 1543
Db 1516 -----EAARGCWYFNSAGQAVTGQOVINGKQLYFDGSGRQVKG 1554
Qy 1544 KAVRTDQKIRYFDENSGSM 1563
Db 1555 RYVYVG-GKRLFCDAKTGEL 1573

RESULT 10
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternanase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499, 203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PR
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match 29.4%; Score 2454; DB 2; Length 2057;
Best Local Similarity 37.4%; Pred. No. 1.6e-163;
Matches 627; Conservative 235; Mismatches 541; Indels 274; Gaps 58;

Qy 35 VASADTDTASDSDSNQAVVTGDTT--NNQATDQTSIAATATSQSASDTAATDQASAAEQ 92
Db 242 VADSSGQTYFDNGQPLGLQLIIDLGMQLVFNQCGVQI-----KGFQDVNNKRIYFAPN 236
Qy 93 TQGTATASD-----TAAQTITTANAEAKWVPTTENENQCGFTDMLAEAKNVAFAESDPSDL 148
Db 297 TGNVAANTEIINGKLQGRDANGQVK-----NAFSKDV--AGNTFFYFDANGV---- 341
Qy 149 AKMSNVKQVDGKYYKYDDODGNVKNFVAVSGDKIYFDP-ETGAYKDTSKVDADKSSAVS 207
Db 342 -MLTGLQTTISGKTYLDEQGHRLKNYAGTFNNOFMFYFDATGAGKTAIEYQFQDGLVSQS 400
Qy 208 QNATIFAANNRAYSTSAKNFEAVDNYLTADSWRPKRSILKDGKTWTESGDKDPRPLLMW 267
Db 401 NENT-...PHNAKSYDKSSFENVVGYLTATWTRPTDILKNGTWTASTETDWRPLMTW 457
Qy 268 WPDTEKRNYYNM-NKVVGIDKTYTAETSQADLTAAAEVLQARI EKITSENNKTWLR 326
Db 458 WPDQIQANYLNFMSKGLGITTTYYTAATSQKTLNDAAFVIQTAIEQIISLKSETEWLRD 517
Qy 327 AISAFAVKTQPOWNGESE-KPYD--DHLQALLFDNQDTLTPDTQS-NYELLNRTPTNQI 382

Db 518 AIDSFVKTOANWKNQTEDEAFGLQWOGFLAYQDDSHRTPTNTDGNRKNLGRQPINID 577
Qy 383 GSLDSRFTYPNPDPIGGYDFLLANDVDSNPVQVQAEQLNWLHYLLNFGSIYANDADANFD 442
Db 578 GSKDT-----TDGKGS-EFLLANDIDNSNPVQVQAEQLNWLHYLLNFGSITGNDNANFD 630
Qy 443 STRVDAVDNVDADLLQISSDYLKAAAGIDGNKNKNANNHVSIVEAWSNDNTPYLLHDDGDL 502
Db 631 GIRVDAVDNVDADLLKIAAGDYFKALYGTDKSDANANKHLSILEDMNGKDPQVYVNOQNAQ 690
Qy 503 MNMKNKFRLSMLSLAKPLDKRS-----LNPLI-----HNSLVDREV 541
Db 691 LTWDYTVTSQFNSLTHGANNRSMWYFLDTGYLLNGDLNKKLVDRNPNSTGLVNRAN 750
Qy 542 DREVETVPYSFARAHSEVDQIIRDIKAEINP---NSFGYSFTQEEIEQAFKIYNEDL 598
Db 751 SGTKVIPNYSFVRAHDYDAQPIR---KAMIDHGIKNMQDTFTFDQLAQGMFEFYKQ 807
Qy 599 KKTDD--KKYTHNVPLSYTLILLTNKGSIPRVYVGDHFTDDGQYMANKTNYDAIESLLKA 656
Db 808 ENPSGFKKYNDYNLPSAYAMLLTNKDTVPRVYVGDHFTDDGQYMANKTNYDAIESLLKA 867
Qy 657 RMKYVSGQAM-----QNYQICNGEILTSVRYGKALKQSDKGDAFT-----RTSG 702
Db 868 RIKYVSGGQTMATDSGKDLKQGETDLLTSVAFGKIM---TSDQTTTQDINSQDYKNOG 923
Qy 703 VGVVMGNQPNFSLDG-KVVALNNGAAHANOYRALMVSTKDGVAITYATDADAKAGLAVR 761
Db 924 IGVIVGNPDKLNDKTTILHMKAKHQLYRALVLSNDSGIDVDYDSDDKAP---TLR 979
Qy 762 TDENGYLYF-----LNDLKGAVNPQVSGFLOVWVPUCAADDQDIR-VAASD 807
Db 980 TNDNGDLIFHKTNTFVKQDGTIINYEMKGLNALISGILGVWVVPVSGASDQARTVATES 1039
Qy 808 TASTDCKSLHODAAAMDSPVMEFGSNFQSFATKEBEYTNVIVANNVDFKFSVSGITDFEMA 867
Db 1040 SSSDGSVFHSAALDSNVIYEGFSNFQAMPTSPQSTNVVIATKANLKFELGITSFELA 1099
Qy 868 PQVVSSTDG-----QFLDSVIONQYAFTRDYDLGMSKAN-----KYGTADQVLVKAIAL 916
Db 1100 PQYRSSGDTNYGMSFLDSFLANGYAFTDYLDFGNKADGNPNPTKYGTQDQDLRNAIEAL 1159
Qy 917 HAKGLKWMADWPDQMYTFPKQEVTVTRTFKFKPIAGSQINHSILYVTDTKSSGDDYQA 976
Db 1160 HKNGMQAIADWVPDQIYALPKGEVVTATRVDERGNQLKDTDFVNLLYVANTKSSGVDYQA 1219
Qy 977 KYGGAFLDELKPKYPELFTKKOISTGOAIDPSVKIKOWSAKYPNGSNILGRGADYVLSQ 1036
Db 1220 KYGGEFLDKREYEPFLFQNGQVSTGQPIDASTKIKOWSAKYNMGNTILHRGAYYVLKDW 1279
Qy 1037 VSNKYFNVA-SDTFLPSSLLGKWBESGIRYDGKGIYINSSATGDQVKASFITEA-GNLY 1094
Db 1280 ATNQYFNIAKTNEVFLPLQKQDAQTGFISDASGVKY-YISGYQAKDTFIEDGNGWY 1338
Qy 1095 YFGKQGYM-----TGAQTINGANYFFLENGTALANTYTDAGNSHYAND 1141
Db 1339 YFDKQGYMVRSGOGENPIRTVETSVNTRNG-NYFMPNGVELRKGFGTDSNGVYYPDDQ 1397
Qy 1142 GKREYNGY--QQFGNDWRYFKDGNMAVGLTVDGN--VOYFDKDGVAQDKKIIVTRDGKVR 1198
Db 1398 GMRVRDKYINDDANNFYHLNDGTVSRGLFKFDSDTLQIFASNGVQIKDSYAKDSKGNKY 1457
Qy 1199 YFDQHNQNAATNFTIADKTGHW-----YILGKDGVAVTGAQTGVGKQKLY--FEANGQOVKG 1252
Db 1458 YFDSATGNNDTG-----KAQTDWNGYIYITIDSDA---NNTIGVNTDYATYITSSLEDG 1509
Qy 1253 DFVTSDEGKLYFYVDVSGDMWTDTFIEDKAGN---WFLYLGKOGAAVTGAQTIRGQKLYFK 1309
Db 1510 LFANAPY-----VVTQDQNGDLNMQY-----INHTKQY-- 1539
Qy 1310 ANGQOVK-----GDIVKGTGDKIRYDAKSGEQVFNKTVKAADGK 1349

Db 1540 -EGQVQVTRQYTSKGVSNLITPAGGDL-----QGRLWVDSRALJTWPFKTNQISFI 1594
Qy 1350 TVYIGNDGAVD-PSVVGQTFKDGAGLRFYNLKGQLVGTGSGWYETANHDWVYIQ----- 1404
Db 1595 SYANRNDGLFLNAPVQVKGYLACNSN-----QYKGQVTTIAGVANVSGKWSLSFNGT 1649
Qy 1405 -----SGKALTGEOT-----INQOHLV--PKEDGHQVKG-----OLVTGT 1437
Db 1650 QYWDSQALNTNFTHDMNQKVFVNTTSLNDGLFLNAPYRQPCYKLAGLAKXNNYNTVTS 1709
Qy 1438 DGKRYRYDANGSDQAFNKS-VTVNGKTYVFGNDGTAQTAGNPKGOTPKGSDIR----- 1490
Db 1710 Q-----QYFD-----DQGTWVSQVVLGGQTVVDNHALAQMVSDTDQQLVYNSNGRNDGLFL 1762
Qy 1491 ---FYSMEGQLVTSQWYENAGQWLYV-KNGKVLTLGLQ-----TVGSRVYFDENGI 1539
Db 1763 NPYRGQGSQLI---GWTADYNGHVQVTKGQDAYGAQWFLITLNNQVQVDSRAL 1816

RESULT 11
US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRF
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

Query Match 27.4%; Score 2284; DB 2; Length 1278;
Best Local Similarity 46.0%; Pred. No. 7.2e-152;
Matches 497; Conservative 167; Mismatches 338; Indels 78; Gaps 29;

Qy 116 VPTENEN-----QGFTEMLAEAKVATAESDSIP-----SDLAKMSNVKQVDGKYY 163
Db 185 VTKNENVLVHRFSNDVKTGEGNYVDFWSELMPVKDSFOKNGPLKQFGLQTINGQOYY 244
Qy 164 YD-QDGNVKNFAVSGDKIYYFD-ETGAYKDTSKVDADKSSAVSONATIPAAANRAYS 221
Db 245 IDTTGGPRKNFLQSGNNWYFDSGTG--GTNALELQFAGKGTVSNEQ-YRNGAAYS 301
Qy 222 TSAKNFEADVNYLTADSRPKSILDKGKTWTSKDDPRFLMAMWPDTEKRYNYVM 281
Db 302 YDDKSIENVNGYLTAADTYRKPQLKDGTTWDSKETDMEPILMVWNPENLTQAYLYNM 361
Qy 282 ----NKVVGIDKTYTAETSQADLTAAELVQARIQKITSNNYKWLREISAIVKTPQ 337
Db 362 KQHGNLLPSALPFNADADPAELNHYSEIVQOQIEKRISSETGTMTDLRLTLMHDFVTNNPM 421
Qy 338 WNESEKPYDD--HLQNGALLFNQDTLTPDTCOSYRLLNRTPTNQTGSLDSRFTYNPND 395
Db 422 WNKDSNNVPSGIFQGGFLKYEN-SDLTPYANSYRLGLRMPINIKQ-----TYR--- 472
Qy 396 PLGGYDFLLANDVNSNPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDVAVNDAD 455
Db 473 ---QGEFLANDIDNSNPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDVAVNDAD 529
Qy 456 LQOISSDYLKAAVGIQDNKNNKANNHYSEIVAEWSDNTPYLLHDDGDLNMNMNFKRLSMLW 515
Db 530 LAMIAQDYFNAAYGMD-SDAVSKHINILEDNMHADPEYFNKICNPQLTMDDTIKNSLNH 598

Qy 516 SLAKPLDKRSLNPLIHNLSVDREVDETVPSPYSFARAHDSVQDIIRDIKABINP 575
Db 589 GLSDATN-RWGLDAIVHQSADRENSTENVVIENYSFVRADNNNSQDIQNAIR-DVTG 646
Qy 576 NSFQYSPQEBEIOAFKIYNEDLKKTKYTHYNPLSYTLTLTKGSIPIRVYIGDMFTD 635
Db 647 KDY-HTTFEDEQKIDAYIQDQNSVTVKYNLYNIPASVAILLTNKDTIPRVYIGDLYTD 705
Qy 636 DGYMANKTVNYDAIESLLKARMKYVSGGQAMQYQI-GNGEILTSVRYGKALKQSKG 694
Db 706 GGOYMEHQTRYDYTLTLNLSRVKYVAGGOSMOTMSVGGNNNLTLSVRYGKAMTATDTG 765
Qy 695 DATRTISGVVGMGNQPNFSL--DGKVVALNMGAHANOEYRALMVSTDCGVATYATDAD 752
Db 766 TDETRTOGIVGVSVNTENLKLGVNDKV-LHMGAAHKNQYRAAVLTITDGVINYSQ 824
Qy 753 ASKAGLVKRTDENGILYFLNDL-----KGVANPQVSGFLQVWVPGAADQD 800
Db 825 AP-----VAMTDENGDLVLSHNLVNGKEADTAVQGVANPDSVGYLAVWVPGASDNQD 880
Qy 801 IRVAASDTASTDGKSLHQDAAMDSRVMPFGFSNFQSFATKEEETVNVVIANVDFVSWG 860
Db 881 ARTAPSTKNSGNSAYRTNAAPDSNVIFEAFSPVYPTTKESERANVRIAQNADFFASLG 940
Qy 861 ITDFEMAPQVSVSTDGQFLDSVIONGYAFTDRYDLGMSKANKYCTADOLVKAIKALHAKG 920
Db 941 FTSFEMAPQVNSKDRFTLSDTIDNGYAFTRDRLGMSKANKYCTADOLVKAIKALHAKG 1000
Qy 921 LKYNADWVPDQMYTFPKQEVYTVTRTDKFGKPIAGSQINHSLYVTDTKSGDDYQAKYGG 980
Db 1001 LQWADWVPDQIYNLPQKEVATVTRVDDRGVNWKDAIINNLYVNT-IGGGEYQKYYG 1059
Qy 981 AFLDELKKEYPETFKQISTQGAIDPSVKIKWSAKYFNGSNILRGADYVLSDDQVSNK 1040
Db 1060 AFLDKLQKYPEIFTKQVSTGVAIDPSQKITMSAKYFNGSNILRGADYVLSDDQVSNK 1118
Qy 1041 YFNVASDT-LFLPSSLLGKVVESGIRY---DGKGYLYNSATGQDVKASFIPEA-GNLY 1094
Db 1119 YNLTGTTTKQFLPIQLTGKQKQNEGFVKNGDNGYYPD--LAGNMVNTFIEDSVGNWY 1176
Qy 1095 YFGKGYMYMTGAQTIN-----GANYFFLENGTALR-NTIYTDAGNNSHYVANDGKRYEN 1147
Db 1177 FPDQDGKGVENKHFVDVDSYGEKGTFFLKNQVSGFRLGVLQTD--NGTYIFDNYGKAVRN 1234

RESULT 12
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRF
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match 27.4%; Score 2284; DB 2; Length 1781;
Best Local Similarity 46.0%; Pred. No. 1.2e-151;


```
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; APPLICANT: LEER, ROBERT-JAN
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-995-749A-10

Query Match      19.5%; Score 1631; DB 2; Length 545;
Best Local Similarity 57.5%; Pred. No. 2e-106;
Matches 315; Conservative 94; Mismatches 117; Indels 22; Gaps 7;

QY 403 LLANDVNSNPVVOAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDADLLQISSD 462
Db 1 LLANDVNSNPVVOAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDADLLQISSD 462
QY 463 YLKAAYGIDKNKNNHVSIVAEWSNDTPYLHDDGDNLMNMDKFRSLMWSLAKPLD 522
Db 61 YLKAHYGVDSKNAHNSILAEWSNDTPYLNKDTKGAQLPDKNLRSLLYALTRPLE 120
QY 523 K-----RSGLNPLHNSLVREDDREVETVPSYFARAHDSVQDIIRDIKAEIN 574
Db 121 KQASNKNEIRSGLEPVITSLNRSAGKNSERMANYIFIRAHDSVQTVIAKIIKQIN 180
QY 575 PMSFGYSFTQEEIQAFKIYNEDLKKTKYTHYNNVPLSYTLTLTKNGSIPRVYGDMP 634
Db 181 PKTDGLTFLDELKQAFKIYNEDMRQAKKKTQSNIPTAYALMNSKDSITRLYGDMS 240
QY 635 DGOYMANITYNDAIESLLKARKMYSGGQAMQ-NYQINGE-----ILTSVRYGK 685
Db 241 DGOYMATKSPYYDAIDTLKARIKAYAGQDMKITVEGDKSHMDWDYTGVLTSVRYGT 300
QY 686 GALKOSDKGATRTSGVGMGNQPNFSLD-GKVVALNMGAAHNOEYRALMVSTKGV 744
Db 301 GANEATDQGEATKTQGMVITSNPNSLKNQNDKRVIVNMGAAHNOEYRPLLLTTKDL 360
QY 745 ATYATDADASKAGLVKRTDENGLYFLNDDLGAV-ANPOVSGFQVWVPVGAADQDIRV 803
Db 361 TSYTSDAAK--SLYRKTNDKGLVFDASDIQGLYNLPQVSG-LAWVPVGSNDQDIRV 417
QY 804 AASDASTDGKSLHODANDSRVMEFGSNFQSFATKEEYTNVVIANNVDFKVSNGITD 863
Db 418 AASNKANATGVYESSSALDSQLIYEGFSNFQDFVTKDSDTYNTKIAQNVQLFKSNGVTS 477
QY 864 FEMAPQVYSTGQFLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLV 923
Db 478 FEMAPQVYSSDGSFLDSIQNGYAFEDRYDLAMSNNKYNKYGSDQDMINAVKALHKSQ 537
QY 924 MADWVPDQ 931
Db 538 IADWVPDQ 545
```

RESULT 15

```
US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
```

```
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: B043388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
; US-09-604-957-5
```

```
Query Match      19.2%; Score 1600; DB 2; Length 523;
Best Local Similarity 58.3%; Pred. No. 2.9e-104;
Matches 311; Conservative 76; Mismatches 132; Indels 14; Gaps 6;

QY 403 LLANDVNSNPVVOAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDADLLQISSD 462
Db 1 LLANDVNSNPVVOAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDADLLQIAAD 60
QY 463 YLKAAYGIDKNKNNHVSIVAEWSNDTPYLHDDGDNLMNMDKFRSLMWSLAKPLD 522
Db 61 YKFLAYGVNDQDANTANQHLSILEDWSHNDPLVYTDGQSQNLTMDDYVHTQLINSLTKSSD 120
QY 523 KSGLNPLHNSLVREDDREVETVPSYFARAHDSVQDIIRDIKAEINP---NSFG 579
Db 121 IRGTMQRFVYDMVDRSNDSTENEAIPTNSYFVRAHDSVQTVIAQIV-SDLYPDVENS- 178
QY 580 YSFTQEEIQAFKIYNEDLKKTKYTHYNNVPLSYTLTLTKNGSIPRVYGDMPDGOY 639
Db 179 -APTEQLAAAFKVYNDEKLDKKTQYNMASAYAMLLTNKDTVPRVYGDLYTDGOY 237
QY 640 MANKTVNYDAIESLLKARKMYSGGQAMQNYQINGEILTSVRYGKALKQSKDGKDATTR 699
Db 238 MATKSPYDAINTLLKARQYVYAGQSM---SVDSNDVLTSVRYGKDATASDTGTSETR 294
QY 700 TSGVGMGNQPNFSL-DGKVVALNMGAAHNOEYRALMVSTKGVVATYATDADASKAGL 758
Db 295 TEGIGVIVSNNAELQLEDGHTVTLHMGAAHKNQAYRALLSTTADGLAYYDTDENAP- 350
QY 759 VKRTDENGLYFLNDDLGAVANPOVSGFQVWVPVGAADQDIRVVAASDTASTDGKSLHQ 818
Db 351 VAYTDANGDLIFTNSIYGVQNPQVSGYLAVWVPVGAQQDQDARTASDTTNTSDKVFHS 410
QY 819 DAAMDSRVMEFGSNFQSFATKEEYTNVVIANNVDFKVSNGITDFEMAPQVYSSTDGOF 878
Db 411 NAAALDSQVIYEGFSNFQAFATDSSSEYTNVVIANNADQFKQGVTSFQLAPQVRSSTDTSF 470
QY 879 LDSVTQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQ 931
Db 471 LDSIIQNGYAFTRDYDLGVTGTPKYGTADQLRDAIKALHASGIQAIADWVPDQ 523
```

Search completed: February 11, 2006, 19:45:54
Job time : 42.8119 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 20:39:23 ; Search time 142.315 Seconds
(without alignments)
4668.162 Million cell updates/sec
Title: US-10-797-821-37
Perfect score: 8349
Sequence: 1 MEKNVRFKMKVKRWVTLUS.....VYQYYFGNDGARIYRGMN 1590
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:*

1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8349	100.0	1590	4	US-10-383-930-37
2	8349	100.0	1590	5	US-10-797-821-37
3	4580	54.9	1475	3	US-09-740-274-2
4	4580	54.9	1475	4	US-10-383-930-34
5	4580	54.9	1475	5	US-10-797-821-34
6	4214.5	50.5	1375	3	US-09-740-274-4
7	4214.5	50.5	1375	4	US-10-383-930-35
8	4214.5	50.5	1375	5	US-10-797-821-35
9	3575	42.8	1430	3	US-09-740-274-6
10	3575	42.8	1430	4	US-10-383-930-36
11	3575	42.8	1430	5	US-10-797-821-36
12	3502.5	42.0	1554	4	US-10-383-930-38
13	3502.5	42.0	1554	5	US-10-797-821-38
14	3325.5	39.8	1518	4	US-10-383-930-40
15	3325.5	39.8	1518	5	US-10-797-821-40
16	3031	36.3	1365	4	US-10-383-930-39
17	3031	36.3	1365	5	US-10-797-821-39
18	2975	35.6	1497	5	US-10-484-218-18
19	2808	33.6	1595	5	US-10-484-218-20
20	2472	29.6	1006	5	US-10-484-218-22
21	2454	29.4	2057	4	US-10-417-280A-2
22	2292.5	27.5	1777	5	US-10-484-218-12
23	2284	27.4	1781	3	US-09-995-749A-2
24	2220.5	26.6	1771	5	US-10-484-218-14
25	1631	19.5	545	3	US-09-995-749A-10
26	1595.5	19.1	522	3	US-09-995-749A-11
27	1401	16.8	535	3	US-09-995-749A-13

28	1368.5	16.4	787	5	US-10-484-218-16	Sequence 16, Appl
29	1260.5	15.1	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1206	14.4	525	5	US-10-484-218-23	Sequence 23, Appl
31	675	8.1	223	5	US-10-484-218-6	Sequence 6, Appl
32	669.5	8.0	224	5	US-10-484-218-4	Sequence 4, Appl
33	653	7.8	223	5	US-10-484-218-10	Sequence 10, Appl
34	553	6.6	221	5	US-10-484-218-2	Sequence 2, Appl
35	543	6.5	221	5	US-10-484-218-8	Sequence 8, Appl
36	514	6.2	2710	4	US-10-011-366-6	Sequence 6, Appl
37	514	6.2	2710	4	US-10-354-774-6	Sequence 6, Appl
38	514	6.2	2710	4	US-10-271-012-6	Sequence 6, Appl
39	514	6.2	2710	4	US-10-729-122-6	Sequence 6, Appl
40	514	6.2	2710	4	US-10-729-039-6	Sequence 6, Appl
41	514	6.2	2710	5	US-10-729-527-6	Sequence 6, Appl
42	514	6.2	2710	5	US-10-727-898-6	Sequence 6, Appl
43	514	6.2	2710	5	US-10-728-696-6	Sequence 6, Appl
44	514	6.2	2710	6	US-11-001-241-6	Sequence 6, Appl
45	459	5.5	866	4	US-10-222-038-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match	100.0%;	Score 8349;	DB 4;	Length 1590;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1590;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MEKNVRFKMKVKRWVTLUSASATMLASALGASVASADTDATDQASAEQQTGTTASTDTAAQTNTTANAEKWPVTEN	60	
Db	1	MEKNVRFKMKVKRWVTLUSASATMLASALGASVASADTDATDQASAEQQTGTTASTDTAAQTNTTANAEKWPVTEN	60	
QY	61	QATDQTSIAATATSEQASASTDAATDQASAEQQTGTTASTDTAAQTNTTANAEKWPVTEN	120	
Db	61	QATDQTSIAATATSEQASASTDAATDQASAEQQTGTTASTDTAAQTNTTANAEKWPVTEN	120	
QY	121	ENQGFTEMLAKGNVATAESDIPSDLAKGNVKQVGGYKYDDQGNVKNFVSVGD	180	
Db	121	ENQGFTEMLAKGNVATAESDIPSDLAKGNVKQVGGYKYDDQGNVKNFVSVGD	180	
QY	181	KIYIFDETGA YKDTSKVDADKSSAVSQNATIPANNRAYSTSAKNFEADVNTLTADSWY	240	
Db	181	KIYIFDETGA YKDTSKVDADKSSAVSQNATIPANNRAYSTSAKNFEADVNTLTADSWY	240	
QY	241	RPKSLKDGKWTESGKDDFRPLLMWPDTEKRYVNMKNVGVGIDKTYTAETSQADL	300	
Db	241	RPKSLKDGKWTESGKDDFRPLLMWPDTEKRYVNMKNVGVGIDKTYTAETSQADL	300	
QY	301	TAAELVQARIPOKITSENNTKWLREAIISAFVKTPQWNGSEKPYDDHLQNGALLFDNQ	360	
Db	301	TAAELVQARIPOKITSENNTKWLREAIISAFVKTPQWNGSEKPYDDHLQNGALLFDNQ	360	

```
QY 361 TDLTPDTQSNRYLLNRTPTNQTSGLDSRFTYNPNDPLGGYDFLLANDVDNSNPVVQAEQL 420
Db 361 TDLTPDTQSNRYLLNRTPTNQTSGLDSRFTYNPNDPLGGYDFLLANDVDNSNPVVQAEQL 420
QY 421 NMLHYLLNFGSIYANDADANFDSIRVDADVNDADLLQISSDYLKAAYGIDKNNKNNNH 480
Db 421 NMLHYLLNFGSIYANDADANFDSIRVDADVNDADLLQISSDYLKAAYGIDKNNKNNNH 480
QY 481 VSIIVEAWSNDPTPLHDDGDNLMNDKFLRLSLMLSLAKPLDKRSGNLPLIHNSLDVREV 540
Db 481 VSIIVEAWSNDPTPLHDDGDNLMNDKFLRLSLMLSLAKPLDKRSGNLPLIHNSLDVREV 540
QY 541 DREVEVTSYSFARAHDSVQDIIRDIKAEINPNSFGYSFTQEBIEQAFKIYNEDLKK 600
Db 541 DREVEVTSYSFARAHDSVQDIIRDIKAEINPNSFGYSFTQEBIEQAFKIYNEDLKK 600
QY 601 TDKKYTHYNVPLSYTLILLTNKSGIPRVYVYGDMPFTDDGQYMAKNTVNYDAIESLLKARMKY 660
Db 601 TDKKYTHYNVPLSYTLILLTNKSGIPRVYVYGDMPFTDDGQYMAKNTVNYDAIESLLKARMKY 660
QY 661 VSGGQAMQYQIENGHILSVRGKALQSKDKGDATRTTSGVGVVMGNQPNFSLDGKV 720
Db 661 VSGGQAMQYQIENGHILSVRGKALQSKDKGDATRTTSGVGVVMGNQPNFSLDGKV 720
QY 721 ALNMGAHAHQEYRALMVSTKGVATYATDADASKAGLVKRTDENGLYFLNDDLGKVN 780
Db 721 ALNMGAHAHQEYRALMVSTKGVATYATDADASKAGLVKRTDENGLYFLNDDLGKVN 780
QY 781 PQVSGFLQVWPVGAADDQDIRVAASDTASTDGKSLHODAAMDSTRVWPEGFSNFQSFATK 840
Db 781 PQVSGFLQVWPVGAADDQDIRVAASDTASTDGKSLHODAAMDSTRVWPEGFSNFQSFATK 840
QY 841 EBEYTNVNLANNVDKPVSWGITDFENAPQVVSSTDGQFLDSVTQNGYAFTRDYDLGMSKA 900
Db 841 EBEYTNVNLANNVDKPVSWGITDFENAPQVVSSTDGQFLDSVTQNGYAFTRDYDLGMSKA 900
QY 901 NKYGTADQLVKAIKALHAKGLKVMADWPDQMTYFPKQEVVTVTRTDKFKPIAGSQINH 960
Db 901 NKYGTADQLVKAIKALHAKGLKVMADWPDQMTYFPKQEVVTVTRTDKFKPIAGSQINH 960
QY 961 SLVYVTDKSGDDYQAKYGAFDELKEKYPELFTKKQISTGQAIDPSPVKIKQWSAKYFN 1020
Db 961 SLVYVTDKSGDDYQAKYGAFDELKEKYPELFTKKQISTGQAIDPSPVKIKQWSAKYFN 1020
QY 1021 GSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLGKVVESGIRVDGKGIYNSSATGD 1080
Db 1021 GSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLGKVVESGIRVDGKGIYNSSATGD 1080
QY 1081 QVKASPFITEAGNLYYFGKDGVMYVTAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140
Db 1081 QVKASPFITEAGNLYYFGKDGVMYVTAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140
QY 1141 DGKRYENGYQOFGNDHRYFKQGNMAGLITVDGNVQYFDKQGVQAKDKIIVTRDQKRVYF 1200
Db 1141 DGKRYENGYQOFGNDHRYFKQGNMAGLITVDGNVQYFDKQGVQAKDKIIVTRDQKRVYF 1200
QY 1201 DQHNGNAAWTFTADKTGHWYLGKDGVAVTGAQTVGKQLYPEANGQQVKGDFVTSDEG 1260
Db 1201 DQHNGNAAWTFTADKTGHWYLGKDGVAVTGAQTVGKQLYPEANGQQVKGDFVTSDEG 1260
QY 1261 KLYFYDVDSGDMWTDFTIEDKAGNWFYLGKDGAAVTGAQTIHQKLYFKANGQQVKGDI 1320
Db 1261 KLYFYDVDSGDMWTDFTIEDKAGNWFYLGKDGAAVTGAQTIHQKLYFKANGQQVKGDI 1320
QY 1321 KGTGDKIRYDDAKSGQVFNKTKAADGKTYVTCNDGVAVDPSVWVKQTFKDGASGALR 1380
Db 1321 KGTGDKIRYDDAKSGQVFNKTKAADGKTYVTCNDGVAVDPSVWVKQTFKDGASGALR 1380
QY 1381 NLKQGLVTSQGWYETANHDWVYIQSGKALTGEOTINGOHLVFKEDGHQVKGQVLTGTDGK 1440
Db 1381 NLKQGLVTSQGWYETANHDWVYIQSGKALTGEOTINGOHLVFKEDGHQVKGQVLTGTDGK 1440
```

```
QY 1441 VRYIDANSQDQAFNKSVTTVNGKTYFFGNDGTAQTAGNPKGQTFKQSGDIRFYSMEQOLVT 1500
Db 1441 VRYIDANSQDQAFNKSVTTVNGKTYFFGNDGTAQTAGNPKGQTFKQSGDIRFYSMEQOLVT 1500
QY 1501 GSGWENAGQOWLYYKNGKVLTLGLQTVGSQRVYFVBENGIOAKGKAVRTSDGKIRYFDENS 1560
Db 1501 GSGWENAGQOWLYYKNGKVLTLGLQTVGSQRVYFVBENGIOAKGKAVRTSDGKIRYFDENS 1560
QY 1561 GSMITNQMKFVYQYFFGNDGARIYRGNW 1590
Db 1561 GSMITNQMKFVYQYFFGNDGARIYRGNW 1590

RESULT 2
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37
```

```
Query Match 100.0%; Score 8349; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKNVRFKHKVKKRWVTLVASATMLASALCASVASADTDATSDSDSNQAVVTGDDTTNN 60
Db 1 MEKNVRFKHKVKKRWVTLVASATMLASALCASVASADTDATSDSDSNQAVVTGDDTTNN 60
QY 61 QATDQTSIAATATSBQASASTDAATQASAAEOTQGTASTDTAAQTNTNANEAKWVPTEN 120
Db 61 QATDQTSIAATATSBQASASTDAATQASAAEOTQGTASTDTAAQTNTNANEAKWVPTEN 120
QY 121 ENQGFTEMLAEAKNVATAESDSIPSDLAKMSNVKQVCGKYYYDQGNVKNFVSVGD 180
Db 121 ENQGFTEMLAEAKNVATAESDSIPSDLAKMSNVKQVCGKYYYDQGNVKNFVSVGD 180
QY 181 KIYFDETCAYKQTSKVDADKSSSAVSQNAATIIPANNRAYSTSAKNFEADVNTLTADSWY 240
Db 181 KIYFDETCAYKQTSKVDADKSSSAVSQNAATIIPANNRAYSTSAKNFEADVNTLTADSWY 240
QY 241 RPKSILKQKWTESGKDDFRPLLMAWPDTTKRNVNMKNVCGIDKTYTAETSOADL 300
Db 241 RPKSILKQKWTESGKDDFRPLLMAWPDTTKRNVNMKNVCGIDKTYTAETSOADL 300
QY 301 TAAAEVLQARIIEQKITSNTKWLREAIISAFVKTOPQWNGESEKPYDDHLQNGALLFDNQ 360
Db 301 TAAAEVLQARIIEQKITSNTKWLREAIISAFVKTOPQWNGESEKPYDDHLQNGALLFDNQ 360
QY 361 TDLTPDTQSNRYLLNRTPTNQTSGLDSRFTYNPNDPLGGYDFLLANDVDNSNPVVQAEQL 420
```

Db 361 TDLTPQTQSNRYLLNTPNTQTSLSRFTYFNDPLGGYDFLLANDVNSPNPVOAEQL 420
QY 421 NMLHYLLNFGSIYANDADANFDSIRVDADVNDVADLLOJSSDYLLKAAAYGIDKNNKANHH 480
Db 421 NMLHYLLNFGSIYANDADANFDSIRVDADVNDVADLLOJSSDYLLKAAAYGIDKNNKANHH 480
QY 481 VSIWEAWSNDTPYLHDDGDNLMNMONKFLRLSMLSLAKPLDKRSGLNPLIHNLSVDREV 540
Db 481 VSIWEAWSNDTPYLHDDGDNLMNMONKFLRLSMLSLAKPLDKRSGLNPLIHNLSVDREV 540
QY 541 DREVEVTPSYSPARAHSEVDODIIRDIKAEINPNPSFGYFTQEBIEQAFKIYNEDLKK 600
Db 541 DREVEVTPSYSPARAHSEVDODIIRDIKAEINPNPSFGYFTQEBIEQAFKIYNEDLKK 600
QY 601 TDKKYTHYNVPLSYTLTLLTNKSGISPRVYVYGDFTDGGQYMWANKTVNYDAIESLLKARMKY 660
Db 601 TDKKYTHYNVPLSYTLTLLTNKSGISPRVYVYGDFTDGGQYMWANKTVNYDAIESLLKARMKY 660
QY 661 VSGGQAMQNYQIENGEBILTSVRYGKALKOSDKGDATRTSGVGVVMGNQPNFSLDGKVV 720
Db 661 VSGGQAMQNYQIENGEBILTSVRYGKALKOSDKGDATRTSGVGVVMGNQPNFSLDGKVV 720
QY 721 ALNMGAAHANOEYRALMWSTKDGVAITYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
Db 721 ALNMGAAHANOEYRALMWSTKDGVAITYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
QY 781 PQVSGFLQWVPVGAADDQDQIRVAASDSTASTGKSLHQAAMDSRVWFGFSGNFQSFATK 840
Db 781 PQVSGFLQWVPVGAADDQDQIRVAASDSTASTGKSLHQAAMDSRVWFGFSGNFQSFATK 840
QY 841 EEEYTNVVIANNVDFKVSNGITDFEMAPQYVSVSTDGFLDSVIQNGYAFTRDYDLGMSKA 900
Db 841 EEEYTNVVIANNVDFKVSNGITDFEMAPQYVSVSTDGFLDSVIQNGYAFTRDYDLGMSKA 900
QY 901 NKYGTADQLVKAIKALHAKGLKMAADWPDQMYTPPKQEBVTVTRTDKFGKPIAGSQINH 960
Db 901 NKYGTADQLVKAIKALHAKGLKMAADWPDQMYTPPKQEBVTVTRTDKFGKPIAGSQINH 960
QY 961 SLVYVTDTKSSGGDYQAKYGGAFLDELKEYPELFTKQISTGOADPSVKIKQWSAKYFN 1020
Db 961 SLVYVTDTKSSGGDYQAKYGGAFLDELKEYPELFTKQISTGOADPSVKIKQWSAKYFN 1020
QY 1021 GSNILGRGADYVLSDOVSNKYFNVASDTLFLPSSLGKVVESGIRYDGKGIYNSSATGD 1080
Db 1021 GSNILGRGADYVLSDOVSNKYFNVASDTLFLPSSLGKVVESGIRYDGKGIYNSSATGD 1080
QY 1081 QVKASFITEAGNLYYFGKDGVMVTGAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140
Db 1081 QVKASFITEAGNLYYFGKDGVMVTGAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140
QY 1141 DGKRYENGYYQFQGNDRWYFKDGNMAVGLTTVDGNVQYFDKGVQAKDKLIVTRDQKVRVF 1200
Db 1141 DGKRYENGYYQFQGNDRWYFKDGNMAVGLTTVDGNVQYFDKGVQAKDKLIVTRDQKVRVF 1200
QY 1201 DOHNGNAATNTFIADKTGHYYLKGXGVAVTGAQTVGKQKLYFEANGQVKGDFVTSDBG 1260
Db 1201 DOHNGNAATNTFIADKTGHYYLKGXGVAVTGAQTVGKQKLYFEANGQVKGDFVTSDBG 1260
QY 1261 KLYFYDVDSGDMWTDFTIEDKAGNPFYLGKOGAAVTGAQIRGQKLYFKANGQVKGDIIV 1320
Db 1261 KLYFYDVDSGDMWTDFTIEDKAGNPFYLGKOGAAVTGAQIRGQKLYFKANGQVKGDIIV 1320
QY 1321 KGTDCGIRYDPAKSGEOPFNKTVKAADGKTYVIIGNDVAVDPSVVKQGTFFKASGALRPY 1380
Db 1321 KGTDCGIRYDPAKSGEOPFNKTVKAADGKTYVIIGNDVAVDPSVVKQGTFFKASGALRPY 1380
QY 1381 NLKGQVLTGSGWYETANHDWYIYQSGKALTGBOITNGOHLYPEKDECHQVKGQVLTGTGK 1440
Db 1381 NLKGQVLTGSGWYETANHDWYIYQSGKALTGBOITNGOHLYPEKDECHQVKGQVLTGTGK 1440
QY 1441 VRYDANSQDQAFNKSVTYVNGKTYFFGNDGTAQTAGNPKGQTFKQGSDFIRFYSMEGQLVT 1500

Db 1441 VRYDANSQDQAFNKSVTYVNGKTYFFGNDGTAQTAGNPKGQTFKQGSDFIRFYSMEGQLVT 1500
QY 1501 GSGWYENAGQWLYVKNQKVLTLGLQTVGSORVYFVDENGIOAKGKAVRTSDGKIRYFDENS 1560
Db 1501 GSGWYENAGQWLYVKNQKVLTLGLQTVGSORVYFVDENGIOAKGKAVRTSDGKIRYFDENS 1560
QY 1561 GSMITNQMFKVYQYVYFNGDARIYRGWN 1590
Db 1561 GSMITNQMFKVYQYVYFNGDARIYRGWN 1590
RESULT 3
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
Query Match 54.9%; Score 4580; DB 3; Length 1475;
Best Local Similarity 56.4%; Pred. No. 3.9e-242;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
QY 1 MEKNVRFKXHKYKRWLTVLSVASATMLASGASVADTDTA-----SDSNQAVVTGQ 56
Db 1 MDKVRVYKLRVYKRWLTVLSVASAVMTLTLLSGGLVKADSNESKSI SNDSNTSVVTANE 60
QY 57 TTN--NQATDQTSIAATSEOSASTDAATDOASAEQQTGTTASTDTTAAQTITNANEAK 114
Db 61 ESNVTEATSEKQEAASSQTNHTVTTSSSTSVVNPKE-----VVSNPYTVGETASNGEKL- 115
QY 115 WYPTENENQGFDEMLABA-----KNVATASDSIPSD-----LAKMSNVKQVDGKY 161
Db 116 -----NQTTVTDKTSBAANNISKQTEADTVDDSNAAQLLEKLPNVEIDGKY 169
QY 162 YYYDQGNVKNFVSGDKIYYFDETGAYKDTSTKVADKSSSAVSQNAITFAANNRAYS 221
Db 170 YYYDNNKVRNTFTLIADGKILHFDETGAYTDTSDTVNK--DIVTTRSNLYKKNQVYD 227
QY 222 TSAKNFEAVDNYLTADSWYRPSILKDGKTWTESKDDPRPLLMWPPDTETKRYNTM 281
Db 228 RSAQSFHVHDYLTAESWYRPSIILKDGKWTQSTEKDFRPLMTWMPDQETQRYNTM 287
QY 282 NKVVGIDKTYTAETSQADLTAAAEVLQVARIKQITSENNTKWLRSAISAFVKTPOWNGE 341
Db 288 NALQGINKTYDDTSNQLQNTAAATIAQIEAKITLKNTDWLRQITSAFVKTSANNSD 347
QY 342 SEKPYDDHLQNGALLFDNQDTDLTPDTQSNRYLLNTPNTQTSLSRFTYFNDPLGGYD 401

Db 348 SEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRLNRTPTNQTGKDPRT--ADNTIGGYE 405
Qy 402 FLLANDVDNSNPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDADLLQISS 461
Db 406 FLLANDVDNSNPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDADLLQIAG 465
Qy 462 DYLLKAAAGTIDKNNKANNHVSIVEAWSDDTPLYLHDDGDNLMNMDKFLSLMLSLAKPL 521
Db 466 DYLLKAAAGTIDKNNKANNHVSIVEAWSDDTPLYLHDDGDNLMNMDKFLSLMLSLAKPL 525
Qy 522 DKRSLNPLIHLNSLVDREVDREVEVTPSYSPARADHSEVDIIRDIKAEINPNPSFGYS 581
Db 526 NORSGNPLITSLNVRTDDNETAAPVSPYSIRADHSEVDLIADIIRKAEINPNVVGYS 585
Qy 582 FTQEEIEQAFKIYNEDLKKTKYTHYNVPLSYTLALLTNKGSIPRYVYGDFTDDQYMA 641
Db 586 FTWEEIKKAFIYNKDLLATEKKYTHYNVPLSYTLALLTNKGSIPRYVYGDFTDDQYMA 645
Qy 642 NKTVNYDAIESLLKARKKYVSGQAMONYQIGNGEILTSVRYGKALKOSDKGDATRTS 701
Db 646 HKTINYEATETLLKARIKYVSGQAMRNOQVGNSEIITSVRYGKALKATDGDRTTTS 705
Qy 702 GGVVVGNOPTSLDGK--VVALNMGAAHANOEVRLMWSTKDGVIATDADASKAGLVK 760
Db 706 GVAVIEGNPURLKASDRVVMNGAHKNQAYRPLLLTDNGIKAYHSDQEA--AGLYR 763
Qy 761 RTDENGILYFLNDLKGAVNPQVSGFLQVVPVGAADDQDIRVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGYANPQVSGYLGVVPVGA--LIRKMFALLRLRPHQOMASVHQ 820
Qy 819 DAMDRSRVPEFSNFOQSFATKEEYTNVIANVVDKFKVSGWIGTDEMAPQVVSDDGOF 878
Db 821 NAALDSRVNPEFSNFOQSFATKEEYTNVIANVVDKFAEWGVTDFEMAPQVVSDDGOF 880
Qy 879 LQSVIONGVAFTDRYDLGKSKANKYGTADOLVKAIKALHAKGLKWMADVPQMYTFPKQ 938
Db 881 LQSVIONGVAFTDRYDLGKSKANKYGTADOLVKAIKALHAKGLKWMADVPQMYTFPKQ 940
Qy 939 EYVTVTRTDKFGPIAGSQINHSLYVTDTKSGDDYQAKYGGAFDELDELKERYPELTKKQ 998
Db 941 EYVTVTRTDKFGPIAGSQINHSLYVTDTKSGDDYQAKYGGAFDELDELKERYPELTKKQ 1000
Qy 999 ISTQQAIDRSVKIKOWSAKYFNGSNILGRADYVLSQVSNKYFVA--SDTLFLPSLL 1056
Db 1001 ISTGVPMDFSVKIKOWSAKYFNGSNILGRADYVLSQVSNKYFVA--SDTLFLPSLL 1060
Qy 1057 GKVBESGIRYDGGYLYNSSATGDQVKASFTIAGNLYYFGKGYVMTGAQTINGANYEF 1116
Db 1061 NQDSQVGFSDGKYVYST--SGYQAKNTFISEGDKWYFDNNGYVMTGAQINGANYEF 1119
Qy 1117 LENGTLALRNTIYTDAGNSHYANDGKRYENGQQF--GNDWRYFKDGNMAVGLTVDGNV 1175
Db 1120 LSNGLQLRAILKNEGDGTVAYNGDGRYENGYYQPMGSRVHRFNNGEMSVGLTVIDGQV 1179
Qy 1176 QYFQDKGVQAKKIIIVTRGKVRYFOHNGNAATNFIADKCHWYLYLKGDKGVAVTGAQT 1235
Db 1180 QYFDEMGYQAKGFVTTADGKIRYFKQSGNMYRNFIENEEGKMWLYLGEDGAATVGSQT 1239
Qy 1236 VGKQKLYFEANGQVKGDFVTSDEGLYFYVDVDSGDMWTDFTIEDKAGNWFYLGKDGAAV 1295
Db 1240 INQGHLYFRANGVQVKGFEVTDHGRISYDGNSGDQIRNRFVRNAQGVFFPDNGYAV 1299
Qy 1296 TGAQTIRGOKLYFKANGQVKGDIVRGTDGKIRYDAKSGEQVFNKTVKAADGKTYVIGN 1355
Db 1300 TGAQTIRGOKLYFKANGQVKGFEVTDHGRISYDGNSGDQIRNRFVRNAQGVFFPDN 1359
Qy 1356 DGVAVDPSVVGQTFKIDAGALRFPYMLKGQLVTGSGWYETANHDWYIYQSGKALTEQTI 1415
Db 1360 NGYAV-----TGARTI 1370
Qy 1416 NGHLYFKEDGHQVKGQLVTGDKVRYVDANSQDQAFNKSVTVNGKTYVFGNDGTAQTA 1475
Db 1371 NGHLYFRANGVQVKGFEVTDHGRISYDGNSGDQIRNRFV----- 1412

Qy 1476 GNPQOTFKDGSIDIRFYSMEGQLVTGSGWYENAOQWLYV-KNGKVLITGLQTVGSQRYVF 1534
Db 1413 -----RNAOQWYFDNNGYAVTGARTINGOHLVF 1442
Qy 1535 DENGIAKGAKAVRTSDGKIRYFDENSWSMI 1564
Db 1443 RANGVQVKGFEVTDHGRISYDANSGERV 1472

RESULT 4

US-10-383-930-34
; Sequence 34, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-34

Query Match 54.9%; Score 4580; DB 4; Length 1475;
Best Local Similarity 56.4%; Pred. No. 3.9e-242;

Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;

Qy 1 MEKVRFRFMHKKVKKWVLSVASATMLASALCASVASADTDTA-----SDSNQAVVTGQ 56
Db 1 MDKVRKYKLRKVKKVVTVSVASAVMTLLTSGGLVKADSNESKQISNDSNTSVVTANE 60
Qy 57 TTN--NOATDPTSIATATSEQSASTDAATQASAAEQTGTASTDTAAQTITNANEAK 114
Db 61 ESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKE-----VVSNPYTVGETASNGEKL- 115
Qy 115 WYPTENENQGTDEMLAEA-----KNVATAESDSIPSD-----LAKMSNVKQVDGKY 161
Db 116 -----QNTTVDKTSSEAAANNISKQTTEADTVDDSNAAANLQILEKLPNVKEIDGKY 169
Qy 162 YYYDQGVNKNFAYSVGDKIYYFDETGAAYKDTSKVDADKSSSAYSONATIFAANNRAYS 221
Db 170 YYYDNGKVRTNFTLIADGKILHFDGTGAYTDTSDITVNMK--DIVTTRSNLYKXNQVVD 227
Qy 222 TSAKNFEADVNTLTADSVWRPKSILKDGKTWTESGKDDPRPLIMAWPOTETKRNVTNM 281
Db 228 RQAQSFHVHDLTBAESWTRPKVILLKDKTWTQSTEKDFRPLMTWPPQETQRYVNM 287
Qy 282 NKVVGIDKTYTATTSQADLTAAAEVLQARIEQKITSENNTKWLREASIPVKTPQWNGE 341
Db 288 NAQLGINKTYDTSNQLNLNIAAATIQAKIEAKITLTKNTDMLRQTIISAFVKVTSANSD 347
Qy 342 SEKPYDDHLQNGALLFDNQTDLTPDTQSNRYLLNRTPTNQTGSLDRSFYNNPDLGGYD 401
Db 348 SEKPFDDHLQNGAVLYDNEGKLTTPYANSNYRLNRTPTNQTGKDPRT--ADNTIGGYE 405
Qy 402 FLLANDVDNSNPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDADLLQISS 461
Db 406 FLLANDVDNSNPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDADLLQIAG 465
Qy 462 DYLLKAAAGTIDKNNKANNHVSIVEAWSDDTPLYLHDDGDNLMNMDKFLSLMLSLAKPL 521
Db 466 DYLLKAAAGTIDKNNKANNHVSIVEAWSDDTPLYLHDDGDNLMNMDKFLSLMLSLAKPL 525


```
QY 522 DKBSGLNPLIHNSLVDREVDREVTVPSSPARAHDSVQDIIRDIKAEINPNPFGYS 581
Db 526 NORSGMNPLTNSLNRDNDNAETAAPVPSYFIRAHDDSEVDLIADIKAEINPNVVGYS 585
QY 582 FTQEEIEQAFKIYNEDLKTKDKYTHYVNPVLSYTLTNTKSGIPRVYVGMFTDDGOYMA 641
Db 586 FTWEEIKAFIEIYNKOLLATEKYTHYNTALSTALLTNTKSSPRVYVGMFTDDGOYMA 645
QY 642 NKTVNYDAIESLARKMYVSGGQAMQNYQINGEILTSVRYGKALKQSDKGDATRTS 701
Db 646 HKTINYEATETLKARIKYVSGGQAMNQVGNSEIITSVRYGKALKATDTGRTTTS 705
QY 702 GUGVVMGNQPNFSLDGK-VVALNMGAAHANOBYRALMVSTKOGVATYATDADASKAGLVK 760
Db 706 GVAVIEGNPSSLKASDRVVMNGAAHKNQAYRPLLLTDDNGIKAYHSDQEA--AGLVR 763
QY 761 RTDENGVLFLNDLKGANPQVSGFLOVWVPVGAADDQDIRVAASDTASTDCK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGANPQVSGYLGVWVPVGA--LTKMFALRLARPHQOQMASVHQ 820
QY 819 DAAMDGRVMEPGFSNFSQSPATKEEYTNVVIANNVDKFSWGITDFEMAPQYVSSTDGQF 878
Db 821 NAALDSRVMEFGSNFQAPATKEEYTNVVIANNVDKFAEWGVTDFEMAPQYVSSTDGSP 880
QY 879 LDSVIONGYAFTDRYDLGHSKANKYGTADQLVKAIKALHAKGLKVMADVPDQMTFPKQ 938
Db 881 LDSVIONGYAFTDRYDLGHSKANKYGTADQLVKAIKALHAKGLKVMADVPDQMTFAPEK 940
QY 939 EVVTVTRTDKFGKPIAGSOLNHSLYVTDTKSSGDDYQAKYGGAPFLDELKEKYPELTKKQ 998
Db 941 EVVTVTRDKYGTGPVAGSQIKNTLYVVDGKSSGQOQAKYGGAPFLDELQAKYPELPARQ 1000
QY 999 ISTGQAIDPSVKIKOWSAKYFNGSNILGRGADYVLSQVSNKYFNVA--SDTLFLPSSL 1056
Db 1001 ISTGVPMDSVKIKOWSAKYFNGTNILGRGAGYVLDQATNYFNISDNKEINFLPKTLL 1060
QY 1057 GKVESGIRYDKGYIYNSATGDQVKASFTTEAGNLYYFGKDGVMVTGAQTINGANYFF 1116
Db 1061 NQDSQVGFSGYDGKGYVYVST-SCYQAKNTFISEGDKWYTFDNNGYMVTGAQSGINGVNYF 1119
QY 1117 LENGTAIRNTIYDAGNSHYANCKRYGVNGVQOF-GNDWRYFKDGNMAVGLTIVDGNV 1175
Db 1120 LENGQLRLAILKNEGTYYVYGNDRRYENGYYQPMGVRHFNNGEMSVGLTIVDGVQ 1179
QY 1176 QYFDKDGVOAKDKIIVTRDGKRYVFDQHNNAATNFIADKTGHWYVYLGKDGVAVTGAQT 1235
Db 1180 QYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNFIEVBEGKWLILGEDGAAVTGSQ 1239
QY 1236 VGKQKLYFRANGQOVKGDFTVSDSEKLYFYVDVSDGDMWTDFTIEDKAGNWFYLGKDGAAV 1295
Db 1240 INGQHLVFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFVRNAQGMFYFDNNGYAV 1299
QY 1296 TGAQITRGOKLYFKANGQOVKGDIVKGTGDKIRYVDKSGEQUFNKTVKAAQKTYVION 1355
Db 1300 TGAARTINGOLLYFRANGVQVKGFEVTDYGRISYYDGNSGDQIRNRFVRNAQGMFYFDN 1359
QY 1356 DGVAVDPSVVKQGTFKDASGALRFYNLKGQLVTGSGWYETANHWDWYIQSGKALTEQTI 1415
Db 1360 NGYAV-----TGARTI 1370
QY 1416 NGQHLVFKBEDGHQVQLVGTGDKVRYDYDANGSGQAFNKSVTVNGKTYFYFGNDGTAQTA 1475
Db 1371 NGQHLVFRANGVQVKGFEVTDHGRISYYDGNSGDQIRNRFV----- 1412
QY 1476 GNPQQTQKDGSDIRFYSNEGQLVTGSGWYENAAQGWLYV-KNGKVLGLQTVGSGQRYVF 1534
Db 1413 -----RNAQGMFYFDNNGYAVTGARTINGQHLV 1442
QY 1535 DENGIOAKGKAVRTSDGKIRYFDENSGSMI 1564
Db 1443 RANGVQVKGFEVTDYGRISYYDANGSERV 1472
```

```
RESULT 5
US-10-797-821-34
; Sequence 34, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-34
```

Query Match 54.9%; Score 4580; DB 5; Length 1475;

Best Local Similarity 56.4%; Pred. No. 3 9e-242; Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;

```
QY 1 MEKNVRFKMHKVKRWVTLVSASATMSALGASVASADTDTA-----SDDSNOAVVTGDQ 56
Db 1 MDKVRVYKLRVKRWVTVSVASAVMTLTLSGGLVKADSNEKSQISNDSNTSVVTANE 60
QY 57 TTN--NQATDQTSIAATATSEQASASTDAATDOASAAEQGTGGTASTDTAAQTITNANEAK 114
Db 61 ESNVTEATSKQEAASSQTNHTVTTSSSTSVVNPKE-----VVSNPYTVGTASNGEKL- 115
QY 115 WYPTENENCGFTDEMLAEA-----KNVATAESDSISPSD-----LAKSNVVKQVDDGY 161
Db 116 -----QNGTTVTDKTESEAAANNISKQTTTEADTVDDSNAAANLQLEKLPNKEIDGKY 169
QY 162 YYYDQDGNVKNFVSVGDKIYYFPDETGAHKDTSKVDADKSSSAVSQNAITFAANNRAYS 221
Db 170 YYYDNNGVKRVNTFLIADGKILHFDGTGAYTDTSIDTVNK--DIVTTRSNLYKKYQVYD 227
QY 222 TSANKFEAVDNYLTADSWRPKSIILKDGKTWTESGKDDFRPLLMAWPDPTETKRNVTVM 281
Db 228 RSAQSFHVHDYLTAESWYRPKYILKDGKTWTQSTEKDFRPLMTWMPDQETQRQYVVM 287
QY 282 NKVVGIDKTYTAETSOADLTAAAEVLQARI EOKITSENNTKWLREAI SAFVKTPQPMNGE 341
Db 288 NAQLGINKTYDTSNQLQNTIAAATIQAIEAKITLKNTDWLRQTI SAFVKTQSAWNSD 347
QY 342 SEKPYDDHLQNGALLFNQDITLPTPTQSNRYLNRTPNTQNTGSLDSRFTYNNPNDPLGGYD 401
Db 348 SEKPFDDHLQNGAVLYDNEGKLTVPYANSRYLRNTPNTQNTGCKDPRYT--ADNTIGGYE 405
QY 402 FLLANDVNSNPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDVADLQLQIS 461
Db 406 FLLANDVNSNPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDADVNDVADLQLQIAG 465
QY 462 DYLKAAYGIDKKNKNNHNSIVBEAWSNDTPLYLHDDGDNLMNMNKPESLMLSLAKPL 521
Db 466 DYLKAAGIHKNDKKAANDHLSILEAWSNDTPLYLHDDGDNLMNMNKPESLMLSLAKPL 525
QY 522 DKRSGNPLIHNSLVDREVDREVTVPSSPARAHDSVQDIIRDIKAEINPNPFGYS 581
```

Db 526 NQBSGHNPLITNSLVNRTDDNAETAAPVSPYSPIRAHDSVQDLADIILKAEINPNVVGYS 585
QY 582 FTQEEIEQAFKIYNEDLKTDKYTHYNVPLSYTLTLTNKGSIPRYVYGMFTDDQYMA 641
Db 586 FTMEIEKKAPEIYNKOLLATEKKYTHYNTALSVALLTNKSVPVRYVYGMFTDDQYMA 645
QY 642 NKTWYDAIESLLKAEKMYVSGQAMONYQIGNGEILTSVRVYKGGALKOSDKGDAATTRS 701
Db 646 HKTINYEAJETLLKARIKYVSGQAMRNQQGVNSEIITSVRVYKGGALKATDGDTRTTS 705
QY 702 GGVVNGNOPNPSLDCK--VVALNMGAAHQEAQVRLMWSTKQGVATYATDADASKAGLVK 760
Db 706 GVAVIEGNNPSRLKASDRVVMNGAAHKQAYRPLLTTDNGIKAYHSDQEA--AGLYR 763
QY 761 RTDENGILYFLNDDLKGVANPQVSGFLQVWPVPGAADDQDIRVAASDTASTDGK--SLHQ 818
Db 764 YTNDRGELIFTAAIDIKGVANPQVSGYLVWPVGAA--LKNFALRLARPHQOMASVHQ 820
QY 819 DAAMDVRMPEGSNFQSFATKEEYTNVVIANNVDKFSVSGITDFEMAPQVSSTDGQF 878
Db 821 NAALDSVRMPEGSNFQSFATKEEYTNVVIANNVDKFAEWGVTDFEMAPQVSSTDGSP 880
QY 879 LDSVIQNGVAFTRDYDLGWSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTFPKQ 938
Db 881 LDSVIQNGVAFTRDYDLGISKNPKYGTADDLVKAIKALHSKGIKVMADWVPDQMYAFPEK 940
QY 939 EVVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKEKPELFTKKQ 998
Db 941 EVVTATRVKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYGGAFLEELQAKYPELFPARKQ 1000
QY 999 ISTGQALDSVKIKOWSAKYFNGSNILGRGADYVILSDQVSNKYFNVA--SDTLFLPSSLL 1056
Db 1001 ISTGVPMDSVKIKOWSAKYFNGTNILGRGAGYVILKDOATNTYFNISDNKEINFLPKTLL 1060
QY 1057 GKVVESGIRYDGHGYTINSSATQDVQKASFITAGNLVYFGKDYGVMTGAOTINGANYFF 1116
Db 1061 NQDSQVGFSDGKGYVYST--SGYQAKNTFISBGDKWYFPDNNGYMTGAQSLNGVNYF 1119
QY 1117 LENGTLALNTIYTDAGNSHYIYANDGKRYENGYYQOF--GNDWRYFKDGNMAVGLTIVDGNV 1175
Db 1120 LSNGLQLRAILAKNEDGTVAIYGNDRYENGYYQFMSGVWRHFNNGEMSVGLTVIDGQV 1179
QY 1176 QYFDDGVQAKDIIIVTROGKVRYFPOHNGNAATNTFIADKTHWYILGKDGVAVTGAQT 1235
Db 1180 QYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYRNFIENEEGKWLVLGSDGAAVTSQT 1239
QY 1236 VGKQKLYFEANGQVKGDFVTSDEGKLYFYVDVDSGDMWTDFTEDKAGNWFYLGKDGAAV 1295
Db 1240 INGQHLVFRANGVQVKGFEVTDHGRISYIDGNSGDQIRNRFVRNAQGWFFYFDNNGYAV 1299
QY 1296 TGAQTIRGQKLYFKANGQVKGDIYVKTGDKIRIYDAKSQEYFNKTVKAAADGKTYVIGN 1355
Db 1300 TGARTINGQLLYFRANGVQVKGFEVTDYGRISYIDGNSGDQIRNRFVRNAQGWFFYFDN 1359
QY 1356 DGVAVDPVVGQTFKDGASGALRYFNLKGQLVTGSGWYETANHDWYVIOGSKKALTGEQTI 1415
Db 1360 NGYAV-----TGARTI 1370
QY 1416 NGQHLVFKEDGHQVKGQVLTGDKVRYVDANSQDAFNKSVTVNGKTYFYFGNDGTAQTA 1475
Db 1371 NGQHLVFRANGVQVKGFEVTDHGRISYIDGNSGDQIRNRFV----- 1412
QY 1476 GNPKGQTFKDGDIRFYSMGBQLVTGSGWYENAGQWLYV--KNGKVLTLGLQTVGSQRVY 1534
Db 1413 -----RNAQGWFFYFDNNGYAVTGARTINGQHLXYF 1442
QY 1535 DENGIOAKGKAVTSDGKIRYFDENGSMI 1564
Db 1443 RANGVQVKGFEVTDYGRISYIDANSGERV 1472

; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match 50.5%; Score 4214.5; DB 3; Length 1375;
Best Local Similarity 60.3%; Pred. No. 3.8e-222;
Matches 811; Conservative 186; Mismatches 296; Indels 53; Gaps 14;
QY 1 MEKNVRFKMHKVKKRWVTLASATMLASALGASVASADTDTASDDSNQAVVTGDOT--T 58
Db 1 MEKKVRFKLRKVKKRWVTVSIASAVVTLTSLSSLVKAD---STDDRQQAVTESQASLVT 57
QY 59 NNOATDQTSIA---ATATSEQASDAATDQASAAEQTOGTTAST--DTRAAQTNTTANEA- 113
Db 58 TSEAAKETLTATDTSTATSATSQPTATVTDVNVSTTNQSTNTTANTANTANFVVKPTTSEQAK 117
QY 114 -----KWVPTENEN---QGPTDEM-----LAAKGNVATAESDS 143
Db 118 TPNSDKIITTSKAVNRLTATGKFPVANNNTAHPKVTVDKIVPIKPKIGKLQKQPSLSQDD 177
QY 144 IPSDLAKMGNVKQDGGYKYVQDGNVKNFVAVSGDKIYVYFDETGAYKDTSKVDADKSS 203
Db 178 IAA-LGNVKNIRKVGKYYKYKEDGTLQKNYALNKGKTFFFDETGALSNNLT-LPSKKN 235
QY 204 SAVSQNATIFAANNRAYTSKKNFEAVDNYLTADSWYRPKSILKDGKTWTESKDDPRPL 263
Db 236 ITNNDNTNSFAQVNVQVSTDVANFEHVDHYLTAESWYRPKYILKDGKWTWTQSTEKDPRPL 295
QY 264 LMAWHPDTEKKNYNNKKNVGVIGDKTYTAETFSQADLTAAAEVQARIQKLTSENNTKW 323
Db 296 LMTWMPDQETQRYNNYNNQAQLGHQTYNTATSPLOLNLAQAQTIQIKBEKITAEKNTNW 355
QY 324 LREASAFVKTOPQWNGESEKPYDDHLQNGALLFDNQDTDLTPDTQSNRYLLNRTPTNQTG 383
Db 356 LKQTLISAFVKTSQAWNSDSEKPFDDHLQKGLLYSNNKLTQANSNYRILNRTPTNQTG 415
QY 384 SLDSRTYNNPNDPLGGYDFLLANDVNSNPVQAEQLNMHLHYLNFSGIYANDADANFDS 443
Db 416 KKDPRTY--ADRTIGGYEFLANDVNSNPVQAEQLNMHLFLMNFNGIYANDPDANFDS 473
QY 444 IRVDAVDNVADLLQISSDYLKAAVGIKKNKNANNHVSIVEAWSNDTPLYLHDDGDNLM 503
Db 474 IRVDAVDNVADLLQIAGDYLKAAKGIHKNDKAANDHLSLLEAWSYNDTPLYLHDDGDNMI 533
QY 504 NMDNKFRLSMLSLAKPLDKRSGINPLIHNLSLVDRVDVDDREVETVPFSYFARAFHDSVQD 563
Db 534 NMDNRLRLSLILSLAKPLNQRSGMNPILITNSLVNRTDDNAETAAPVSPYSPIRAHDSVQD 593

Db 952 VVADWVPDQMYALPEKEVVTATRVKIGTPVAGSQIKNTLYVVDGKSGKQDQAKYGGAF 1011
QY 983 LDELKEKYPBELFTKKQISTQQAIDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYF 1042
Db 1012 LEELOAKYPBELFARKQISTGVPMPSVKIKQWSAKYFNGTNILGRGAGYVLDQATNTYF 1071
QY 1043 NVASDTLFLPSSLLG-----KVVESGIRYDGKGYIYNSSATGQVQKASFTTEAGNLYYFG 1097
Db 1072 SLVSDNTFLPKSLVNPVNHGTSSTVTLGVFDGKGYVYST-SGNQAKNAFISLGNWYFYD 1130
QY 1098 KQGYMVTGAQTINGANYFLENGTALRNTLYTDAQNSHYVYANDGKRYENGQOQFGNDWR 1157
Db 1131 NNGYMTGAQSIGNANYFSLNGIQRNALYDNGKVLSTYNGDGRRYENGYLYFGQQR 1190
QY 1158 YFKDGNMAGVLTVDGNVOYFDDKGVQAKDKIIVTRDGVRYPDQRNGNAATNTFIADKT 1217
Db 1191 YFQNGIMAVGLTRVHGAVQYFDASGQAKGQFITTADGKLYRFDKRGQISNRFVRSK 1250
QY 1218 GHYYLKGQGVAVTGAQTVGKQKLYFEANGQQVKGDFVTSDEKLYFYVDVDSGDMWTDTF 1277
Db 1251 GEWFLPDHNGVAVTGTVTENGQRLYKPNQVQAKGEFIRDANGYLYRYDPNSGNEVRNF 1310
QY 1278 IEDKAGNWEVLKDGGAQVGAQTIRG 1303
Db 1311 VRNSKGWELFDHNGIAVTGARVNG 1336

RESULT 8

US-10-797-821-35
; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PR1
; ORGANISM: Streptococcus mutans
US-10-797-821-35

Query Match 50.5%; Score 4214.5; DB 5; Length 1375;
Best Local Similarity 60.3%; Pred. No. 3.8e-222;
Matches 811; Conservative 186; Mismatches 296; Indels 53; Gaps 14;
QY 1 MEKNVRFKMKVKRWVTLVSASATMLASALGASVASADTDTASDDSNQAVVTGDOT--T 58
Db 1 MEKNVRFKMKVKRWVTVSIAAVVTLTSLSGSLVKAD---STDDRQQAATVTSQASLVT 57
QY 59 NNOATDQTSIA---ATATSSQASASTDAATDOASAEQTOGTTAST-DTAAQTNTNANE- 113
Db 58 TSEAAKETLTATDTSTATSATSQPTATVTDNVSTNQSTNTTANTANTANFVVKPTTTSEQAK 117
QY 114 -----KVVPTENEN---QGFTDEM-----LAAEAKNVAETSDS 143

Db 1118 TDNSDKIITTSKAVNRLTATGKFFVPANNNTAHPKVTVDKI VPIKPKIKGLKQPSLSQDD 177
QY 144 IPSDLAKNSNVKQDKYVYDQDGNVKNFVSGDKIYYFDETGAYKDTOSKVDADKSS 203
Db 178 IAA-LGNVKNIRKVGKYYKEDGTLQKNYALNKGKTFPFDGTGALSNN-T-LPSKGN 235
QY 204 SAVSQNATIFAANRAYSTSAAKNFEAVDNYLTADSWYRPKSLKDGKWTESGKDDFRPL 263
Db 236 ITNNNTSFAQYQVISTDVANFBEHVDHLYTAESWYRPKYILKDGKWTQSTEDKFRPL 295
QY 264 LMAWPDTEKKNYVNMKVGVIDKTYTAETSQADLTAAAEVLQARIKQKITSNNTKW 323
Db 296 LMTWPDQETQYVNMNAQIGIHQYNTATSPLOLNLAQTIQTKIEKITAENKTNW 355
QY 324 LREAI SAFVKTQPNWGESEKPYDHLQNGALLFNQDTLTPDPTOSNYLLNRTPTNQG 383
Db 356 LQOTISAFVKTQSAWNSDEKPFDDHLQKGLLYSNNSKLTQANSNYRLNRTPTNQG 415
QY 384 SLDSRFTYNDPDLGGYDELLANDVDSNPVQASQOLNWLHYLLNFGSIYANDADANFDS 443
Db 416 KKDPRT--ADRTIGGEYFELLANDVDSNPVQASQOLNWLHFLMNFENIYANDPDANFDS 473
QY 444 IRVDAVDNVDADLLOISDYLKAAAYGIDKKNKNNHVSIVEAWSNDPTPYLHDDGDNLN 503
Db 474 IRVDAVDNVDADLLOIAGDYLKAAGIKHNDKKAANDHLSILEAWSYNDTPYLHDDGDNMI 533
QY 504 NMDNKRFLSMLWSLAKPLDKRSGLNPLIHNSLVDRVDRVETVPSYSPARAHSEVQD 563
Db 534 NMDNRLRLSLYLAKPLNQRSGMNPITNLSLVNRDNDNAETAAVPSYSPARAHSEVQD 593
QY 564 ITRDIKABINPNSFGYFTQBEIQAQFIYNEDLKKTDKKTYHNVPLSYTLNLLNKGK 623
Db 594 LIRNIIRTEINPNVVGYSFTBEIKKAFIYNKDLLEATKKYTHYNTALSYALLNKGK 653
QY 624 IPRVYVGMFTDQGYMANKTVNYDAIESLLKARKMYVSGQAMONYQICNGEILTSVRY 683
Db 654 VPRVYVGMFTDQGYMAHKTINYEALTELLKARIKYVSGGQAMRQVNSIITSVRY 713
QY 684 GKALKQSDKGDATRTTSGVVMGNQPNFSLDGK-VVALNMGAAHANOEYRALMYSTKD 742
Db 714 GKALKATDGTDRTRTSGVAVIEGNPNSLRLKASDRVVVNMGAHKQAYRPLLTNDN 773
QY 743 GYATYATDADAKAGLVKXTDENGILYFLNDDLKGVANPQVSGFLOVWPVGAADQDIDR 802
Db 774 GIKVHSDQEA--AGLVRYTNDRGELIFTAADIKGVANPQVSGYLVGWWPVGAAADQDVR 831
QY 803 VAASTASTDGKSLHODAAAMDSRVMPGFSNFQSFATKEEYTNVVIANNVDFKFSWGIT 862
Db 832 VAASTAPSTDGKSVHONALDSRVMPGFSNFQSFATKEEYTNVVIANNVDFKFAEWGT 891
QY 863 DFEMAPQYVSSTDGQFDSVIQNGYAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLK 922
Db 892 DFEMAPQYVSSTDGSPDLSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHAKGLK 951
QY 923 VMADWVPDQMYTFPKQEVVTVTRTDKFGKPIAGSQINHSILYVTDTKSSGDDYQAKYGGAF 982
Db 952 VMADWVPDQMYALPEKEVVTATRVKIGTPVAGSQIKNTLYVVDGKSGKQDQAKYGGAF 1011
QY 983 LDELKEKYPBELFTKKQISTGQAIDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYF 1042
Db 1012 LEELOAKYPBELFARKQISTGVPMPSVKIKQWSAKYFNGTNILGRGAGYVLDQATNTYF 1071
QY 1043 NVASDTLFLPSSLLG-----KVVESGIRYDGKGYIYNSSATGQVQKASFTTEAGNLYYFG 1097
Db 1072 SLVSDNTFLPKSLVNPVNHGTSSTVTLGVFDGKGYVYST-SGNQAKNAFISLGNWYFYD 1130
QY 1098 KQGYMVTGAQTINGANYFLENGTALRNTLYTDAQNSHYVYANDGKRYENGQOQFGNDWR 1157
Db 1131 NNGYMTGAQSIGNANYFSLNGIQRNALYDNGKVLSTYNGDGRRYENGYLYFGQQR 1190
QY 1158 YFKDGNMAGVLTVDGNVOYFDDKGVQAKDKIIVTRDGVRYPDQRNGNAATNTFIADKT 1217
Db 1191 YFQNGIMAVGLTRVHGAVQYFDASGQAKGQFITTADGKLYRFDKRGQISNRFVRSK 1250


```
RESULT 10
US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-36

Query Match          1      42.8%; Score 3575; DB 4; Length 1430;
Best Local Similarity 48.4%; Pred. No. 4.5e-187;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

QY 1 MEKNVFRKMHKVRWVTLVSASA--TMLASALGASVAS-----ADTDITAS 44
DB 1 METKRRYKMHKVKHWWTVAVASGLITLGTTLGSSVSAETEQTSDKVVTQKSEDDKAA 60
QY 45 DDSNQA---VVTGDQTTNNQATDQTSIAATATS-EQASASTDAATDQASAAEQQTGTAST 100
DB 61 SESSQTDAPKTQKAQTEQTOAQSOANVADTSTSIKETPSONITQANSDDKTVTNKTSE 120
QY 101 D--TAQATTNANEAKVPTENENQGFTEMLAEAKNVAT-AESDSIPSLAKMSNVKQV 157
DB 121 EAQTSERTKQSEAQ---TFASSQALTOAKBELTKQRTAAQENKPNVDLAAIPNVKQI 177
QY 158 DGKYYYYDQGNVKNFVAVSGVKIYYFDE-TCAYKDTSKVDADKSSVASQONATIFAAN 216
DB 178 DGKYYIGSDQPKKPFALTAVNKVLYFDKVTQALTDTSQYQKQGLTKLND--YTPH 234
QY 217 NRAYSTSAKNFEAVDNYLTADSWYRPSKILKDGKTWTEGKODRPLLMAMWPDPTTKRN 276
DB 235 NQIVNFENTSLETIDNVVTADSWVRPKDILKNGKTWTASSESDLRPLMSWDPDKQTQIA 294
QY 277 YVNYMKV-VGIDKTYTAETSOADLTAAAEVLQARIEQKITSENNTKWLBEAISAFVKYQ 335
DB 295 YLYNMNQQLGIGTGENYTAADSSQSLNLAQTVQVKIETKISQQTQQLRDLIINSFVKYQ 354
QY 336 POWNGSEKPYD---DHLQNGALLFDNQTDLPTDQSNRYLLNRPTNQTGSLDSRFTY 391
DB 355 PNWNSQTESDTSAGEKDHLOGGALLVSN-SDKTAYANSDYRLNLRPTTSQTG----KPKY 409
QY 392 NPNDPLGGYDFLLANDVDSNPNVVAEQQLNWLHYLLNFGSIYANDADANPDSIRVDAVDN 451
DB 410 FEDNSSGGYDFLLANDIDNSNPVVAEQQLNWLHYLMNYGSIIVANDPEANFDGVRVDAVDN 469
QY 452 VDADLIQISDYLKAAAGIDKNNKNNHNSVIVBANSNDNTPYLHDDGDMNLMNMNDKFL 511
DB 470 VNADLLQIASDYLKAHYGVDDKSEKNAINHLUSILEAWSNDNPQYNKDTKGQQLPIDNKLRL 529
QY 512 SMLWSLAKPUDK-----RSLGNPLHNSLVDREVDREVTPVSYPSPARAHDSVQD 563
DB 530 SLIYALTRPLEKASNKWEIRSGLEPVITNSLNRSAEGKNSRMANYIPIRAHDSVQT 589
QY 564 IIRDIITKAEINPNSFGVSTQEBIEQAPKFIYNEDLKKTKDKKTYHNVPLSYTLLTNKGS 623
DB 590 VIAKIIKAQINPKTDGLTFLDELKQAFKIYNEDMRQAKKKYQTSNIPTAYALMLSNKDS 649
QY 624 IPRVYVGMFTDDQYQYANKTVNYDAIESLLKARKMKVSGGQAMQ-NYQIGNGE----- 676

Db 650 ITRLYGDMYSDGQYMATKSPYYDAIDTLKARKIYAAGGQDMKITYYEGDKSHMDWDY 709
QY 677 --ILTSVRYGKALKQSKDGAATRTSGVGVVMGNQPNFSLD-GKVVALNMGAAHANQEY 733
DB 710 TGVLTISVRYGTCAEATDQGEATKTQGMVAVITSSNPSLKLQNQDKVIVNMGAAHKNQEY 769
QY 734 RALMVSTKDGCVATYATDADASKAGLVKRTDENGILYFLNDDILKGVANPOVSGFLQVWV 793
DB 770 RPELLTTKDGLTSTYTSDAAK--SLYRKTNDKGELVDFDASDIQGYLNPQVSGYLAVWV 827
QY 794 GAADDQDIRVAASDTASTDGLSHLQDAAMDSVRMVEGFSNFSQSFATKEEYTNVVIANNV 853
DB 828 GASDNQDVRVAASNKANATGQVYESSALDSQIIYEGFSNFDQFVTKDSYTNKKIAQNV 887
QY 854 DKFVSGIITDFEMAPQYVSSTDGQFLDSVIQNGYAFPTDRYDLGMSKANKYGTADQVLKAI 913
DB 888 QLFKSGVTSFEMAPQYVSSSDGSLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAV 947
QY 914 KALHAKGLKVMADWVPDQMYTFPKQEVVTVTRTDKPKFIAGSQINHSILYVTDTKSSGD 973
DB 948 KALHKSIGIQUIADWVPDQIYNLPKQEVVTVATRVNDYGEYRKDSEIKNTLYAANTKSGKD 1007
QY 974 YQAKYGGAPLDELKEKYPELFTKKQISTGOAIDPSVKIKQWAKYFNGSNILGRGADYYVL 1033
DB 1008 YQAKYGGAPLSELAAYKYSIFNRTQISNGKKIDPSEKITAWAKAYFNGTNILGRGVYVL 1067
QY 1034 SDQVSNKYFNVASDTLFLPSSLLGKVVEGIRYDGKGYIYNSSATGDQVKASFITPA-GN 1092
DB 1068 KDNASDKYPELKGNQTYLPKQMTNKEASTGFVNDGNGMTFYST-SCYQAKNSFVQDAKGN 1126
QY 1093 LYVFGKDGWMTGAQTINGANYFFLENGTALRNTIYTDAGQNSHYIYANDKRYENGYQOF 1152
DB 1127 WYTFDNNGHMYTGLQNGEVQYFLSNGVQLRESFLENADGSKNYEPHGLNRYSNGYYSF 1186
QY 1153 GND--WRYF-KDGNMAVGLTTVDGNVYFDPKDGVOAKDKIIVTRDQKVYFDQHNNAAT 1209
DB 1187 DNDKRWYFDASGVAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKYFDDGSGNMAV 1246
QY 1210 NTFIADKTHWYLYLKGDVAVTGAQTVGKQKLYFEANGQVKGDFVTSDEGLYFYDVIDS 1269
DB 1247 NRFANDKNGDWYLYNSDGLVGVQTINGKTYVFGQDQKQIKGKIIT-DNGKLYFLANS 1305
QY 1270 GDMWTDFTFEDKAGNWFYLGKGAATGAQTTGKQKLYFKANGQVKGDIIVKGTDKIRY 1329
DB 1306 GELARNIFATDSQNNWYFSGDVAVTGQTIAKKLYFASDQKQVKSFTV-YNGKVHY 1364
QY 1330 YDAKSGEVPNTKVAADGKTVYVIGNDGVAVDPSVVKGTQFKDASGALRYNLKGLQVLTG 1389
DB 1365 YHDSGELQVNRFEADKQG----- 1383
QY 1390 SGWYETANHDMWYVYIOS-KGALTGEBQTINGQHLKYFKBDGHQVKGQL 1433
DB 1384 -----NWYLYDSNGEALTGSQRINDQRVFFTTREGKQVKGDV 1419

RESULT 11
US-10-797-821-36
; Sequence 36, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
```


;; PRIOR APPLICATION NUMBER: 09/290,049
;; PRIOR FILING DATE: 1999-04-12
;; PRIOR APPLICATION NUMBER: 60/081,550
;; PRIOR FILING DATE: 1998-04-13
;; PRIOR APPLICATION NUMBER: 60/115,142
;; PRIOR FILING DATE: 1999-01-08
;; NUMBER OF SEQ ID NOS: 45
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 36
;; LENGTH: 1430
;; TYPE: PRT
;; ORGANISM: Streptococcus mutans
US-10-797-821-36

Query Match 42.8%; Score 3575; DB 5; Length 1430;
Best Local Similarity 48.4%; Pred. No. 4.5e-187;
Matches 719; Conservative 240; Mismatches 408; Indels 118; Gaps 26;

QY 1 MEKNVREKMKVKRWVTLVSASA--TWLASALCASVAS-----ADDTAS 44
DB 1 MEKRRYKMKVKRWVTVAVASGLITLGTTLGSSVSAETEQQTSKVVTKSEDDKAA 60
QY 45 DDSNQA---VVTGDTNNQATDTSIAATATS--EQSASTDAATDOASAAEOTGTTAST 100
DB 61 SSSQTDAPTKQAQTEQTAQSQANVADTSTSTKETPSQNIITQANSDDKTVNTKSE 120
QY 101 D--TAAQTNNANEAKVPTTENENQGFTEMLAEAKNVAT--AESDSIPSLAKMSNVKQV 157
DB 121 EAQTSERTKQSEBAQ---TTASSQALTQAKAELTKQRTAAQENKPNVDLAALPNVKOI 177
QY 158 DGKYYVYDQGNVKQPNFVSGVKIYYFDE--TGAYKDTSKVDADKSSASVQNAIPAA 216
DB 178 DGKYYVYIGSDQPKQNFALVNNKVLVFDKNTGALTDTQYQFKQGLTKLND---YTPH 234
QY 217 NRAYSTSANKFEAVDNYLTADSNVRPKSILKQKWTWESKDDFRPLIMAWWPDYTKRN 276
DB 235 NQIVNENTSLETIDNVTADSNVRPKDILKNGKWTWASSESDLRLPLMSWFPDKQTQIA 294
QY 277 YVNNMKV--VGIDKTYTAETSQADTAAAEVLQVARIQKITSNNTKWLREATSAFVKTO 335
DB 295 YLNNMQOGLGTGENTYADSSQSLNLAQTVQVKIETKISQTOQWLRLDINSFVKTO 354
QY 336 PQWNGSESEKPYD---DHQNGALLFDNQDRLTPTQSNVRLNRTPTNQTSGLDSRFTY 391
DB 355 PNWNSQTESDTSAGEKDHQGGALLYSN--SDKTAVANSYRLLNRTPTSTQ---KPKY 409
QY 392 NPNDPLGGYDFLLANDVNSNPVVAEQLNWLHLYLNFSGIYANDADANFDSIRVDADV 451
DB 410 FEDNSSGGYDFLLANDIDNSNPVVAEQLNWLHLYLNYGSIIVANDPEANFDGVRVDADV 469
QY 452 VDADLLQISDYLLKAAAYGDKNNKANNHVSIVEAWSNDNTPYLHDDGDMNMMNKPRL 511
DB 470 VNADLLQIASDYLLKAHYGVDSKSEKALNHLISLEAWSNDNTPQNTKTKGALPDKRL 529
QY 512 SMLWSLAKPLDK-----RGLNPLHNSLVDRVEDVETVPSYFARAHSEVQD 563
DB 530 SLLYALTRPLEKDAOSKNEIRSGLEPVITNSLNNRSAGKNSERMANYIFIRAHSEVQT 589
QY 564 IIRDIKAINPNSFGYSETOEBIEQAFKYNEDLKKTKYHYNVPLSYLLLNKGS 623
DB 590 VIAKIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKTKYQSNIPYAYALMLSNKDS 649
QY 624 IPRVYVYGMFTDQGYMANKTVNYDAIESLLKARMKYVSGQAMQ--NYQINGE----- 676
DB 650 ITRLYYGDMSYDGGQYMATKSPYDAIDTLKARKIYAAGGQDMKITYVEGDKSHMDWDY 709
QY 677 --ILTSVRYGKALKQSKDGAATRTTSVGVVMGNQPNFSLD--GKVVALNMGAAHANQY 733
DB 710 TGVLTSVRYGTGANEATDQGESEATKTQGMNAVITSNPNPSLKLQNDKVKVYNMGAHKNQY 769
QY 734 RALMVSTKGVATYATDAASAKAGLVKRTDENGLYFLNDDLLKGVANPOVSGFLQVWVPV 793
DB 770 RPLLTTKDGLTSYTSDAAK--SLYRKTNDKGLVFDASDIQGYLNPQVSGVLAWVWVPV 827

QY 794 GAADDQDIRVAASDTASTDGKSLHQDAAMDSRVMEFEGSFQSFATKEEETVNVVIANV 853
DB 828 GASDNQDVRVAASNKANATQVYESSALDSQLIYEGSFQDFTKDSDVTNKKIAQNV 887
QY 854 DKFVSGITDFEMAPQVSVSTDGQFLDSVIONGVAFTRDYDLGMSKANKYGTADQLVKAI 913
DB 888 QLFKSGVTSFEMAPQVSVSSDGSFLDSIIQNGVAFEDRYDLAMSNNKYSQQDMINAV 947
QY 914 KALHAKGLKYMADWVPDQMYTFFKQEVVTVTRTDKFKPKTAGSINHSIYVTDTKSSGDD 973
DB 948 KALHKSIGIADWVPDQIYNLPKQEVVTVTRVNDYGEYRKDSSEIKNTLYAANTKSGND 1007
QY 974 YQAKYGGAFDLDELKEKYPELFTKKQISTGQAIIPDSVVIKOWSAKYFNGSNILGRGADYVL 1033
DB 1008 YQAKYGGAFSELAELAAKYPISFNRTQISNGKKIDPSEKITAUKAKYFNGTNILGRGVYVL 1067
QY 1034 SDQVSNKYFNVAADTFLPSSLLGKVVESGIRVDGKGYIYNSSATGQOVKASFITEA--GN 1092
DB 1068 KDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDGNGMTFYST--SGYQAKNSFVQDAKGN 1126
QY 1093 LYTFGKDGVMVMTCAQTINGANYFFLENGTALRTIYTDAGNSHYIYANDGKRYENGQOQF 1152
DB 1127 WYTFDNNGHMVYGLQQLNGEVQYFLSNGVOLRESFLENADGSKNYFGLHGNRYNSGYYSF 1186
QY 1153 GND--WRYF--KDCGNMAVGLTVDGNVQYFDKQVQAKDKIIIVTRDQKRVYFDQHNNAAT 1209
DB 1187 DNDKWRYPDASGVMAVGLKTINGNTQYFDQDGYQVKGAVITGSDGKKRFPDGGSGNNAV 1246
QY 1210 NTFIADKTHWYLLGKDGVAVTGAQTVGKOKLYFEANGQQVKGDFVTSDEGLKYFYDQVDS 1269
DB 1247 NRPANDKNGDWYLLNSDGIALVGQVQTINGKTYFYFGDGKQIKGKIIT--DNGKLKYFLANS 1305
QY 1270 GDMWTDFTEDKAGNFWYLGKQCAAVTGAQTIRGOKLYFKANGQQVKGDIIVKGTGDKIRY 1329
DB 1306 GELARINIFATDSQNNWYFYGSDGVAVTGSGQTIAGKLYFASDGKQVKGSGFT--YNGKVHY 1364
QY 1330 YDAKSGEQVFNKTVKAADGKTYVIGNDGVAVDPSVVKGTFFKDGASGALRFVNLKQLVTG 1389
DB 1365 YHADSSELQVNRPEADKG----- 1393
QY 1390 SGWYETANHDWVYIQS--GKALTGEQTINGQHLFYKEDGHQVKGQL 1433
DB 1384 -----NWYLDNSGEALTGSGRINDQRVFTREGKQVKGDV 1419

RESULT 12
US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38

Query Match 42.0%; Score 3502.5; DB 4; Length 1554;
Best Local Similarity 45.4%; Pred. No. 4.7e-183;
Matches 749; Conservative 252; Mismatches 482; Indels 167; Gaps 40;

QY 47 ----SNOAVVTGDTNNQATD-----OTSIAATAT-----SEOSASTDAATDOASAA 90
Db 60 SLVTSPEATKBAKRTNTKEADVLTTPAKETNAVETATTTNTQATAEAAATTATTADVAVAA 119
QY 91 EQTOGTAGTAAOATTNANE-AKWVPTENENOGFTDMLAEAKNVAEASDPSDLA 149
Db 120 VPKNEAVTTDPAVTTKEAEQOPATVKAENV-----TEVKAPEAALKDSEVEAAL 171
QY 150 KMSNVKQVDKYYYDQDGNVKNFAVSVDGIYYFEDETGAOKDTSKVDADKSSSAVSQN 209
Db 172 SLKNIKNIDGKYYVNEGSHKENFAITVNGQLLYFGKDGLTSSSTYSFTPTGTTNVDG 231
QY 210 ATIFAAANRAYSASKAFPAVDNYLTADSWYRPKSLKDKGTWETSGKDDFRPLLMWWP 269
Db 232 ---FSINNRAYDSSEASFELIDGYLTADSWYRPASIKDGVTAQSTAEEDFRPLLMWWP 288
QY 270 DTETKENYVNMKVVGIDKTYTAETSQADLTAAELVQARTKOTITSENNTKWLREALS 329
Db 289 NVDTVQVLYNYSKVFNLDAKYSSTDQETLVKAAKDIQIKLEQKLOAEKSTQWLRETS 348
QY 330 AFVKTOPQNGSESEKPY-----DDHLQNGALLFDNQDLDLTPDQSNRYLLNRTPTNQTGS 384
Db 349 AFVKTOPQNKETEN-YSKGGGEDHLQGGALLVYVND-SRTPWANSDYRRLNRTATNQTGT 406
QY 385 LBSRFTYNDP--LGGYDFLLANDVNSNPVQAOBQALWHLVYLLNFGSIYANDADANFD 442
Db 407 IDKSIILDEQSDPNHMGDFLLANDVLSNPVQAOBQALNQHLYLMNWSIVMGDKDANFD 466
QY 443 STRVADVNDVADLLQISSDYLKAAAGIDKNNKANNHYSIVAEWSDNDTPYLLHDDGDL 502
Db 467 GLRVADVNDVADMLQVYTFREYGVNKSSEANLAHLSVLEAWSLNDHNNDKTDGAA 526
QY 503 MNMDKFRLSMLSLAKPLDKRS-GLNPLIHNLSLVREVDNR-----EVEVTV 548
Db 527 LAMENKQRLALLFLAKPIKERTPAVSPLYNNTFNTTQORDEKTDWINKDGSKAYNEDGTV 586
QY 549 P-----SYSPARAHSEVODIIRDIKABINPNSFGYSFTQBEIEQAKF 592
Db 587 KOSTIGKYNEKYGDASGNVYFIRAHNNVQDIIAETIKKEINPKSDGFTITDAEMKQAF 646
QY 593 IYNEDLKKTDKKTYHNPVLSYLLLTNKGSI PRVYVYGMFTDDGGYMAKNTVNYDAIES 652
Db 647 IYNKMLSDKKTYLLNPIPAAYVLMQNMETITRVYVYGDYTDGHHMETKSPYYTIVN 706
QY 653 LLKARMKYVSGGQAMQNYQI-----GNGEILTSVRYGKALQKSD-KGDAT 697
Db 707 LMSRKIYVSGGQARQSYWLPDQKMDNSVLYRTNEVVTSVRYGKDIWANTDEGSKY 766
QY 698 TRTSGVGVVGNQPNFSLDGKV-VALNMGAAHANQBYRALMVSTKDGCVATYATDADASKA 756
Db 767 SRTSGQVTLVANNPKLNDQSAKLNVEMGKIHANQYRALIVGTADGKINFTSDADAIAA 826
QY 757 GLVKRTDENGILYFLNDLLKGVANPQVQLQVWVPVGGADDDOIRVAASDASTDGK-S 815
Db 827 GYVKETDSNGVLTFGANDIKGYETTFDMSGEVAVVVPVGGASDNODIRVAPSTEAKKEGELT 886
QY 816 LHQDAAMDVRMPEGFESNFQSF--ATKEEYTVNVIANNVDKFSVWGITDFENAPQVSS 873
Db 887 LKATEAYDSQLIYEGFSNFQIPDGDSPSVYTNKIAENVDFKSNGVTSFEMAPQFVSA 946
QY 874 TDGQFQDSVITQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMY 933
Db 947 DDGTFQDSVITQNGYAFADRYDLAMSNNKYSGSKEDRLDALKALHAKGIAQIADWVPDQY 1006
QY 934 TPKQEVVTVTRTDKFGKPTAGSOIHNLSLVVTDTKSSGDDYQAKYGAFIDELKEKYPEL 993
Db 1007 QLEPEKEVTVTRTDGAGRKIADAILIHSLLVANSKSGKDYQAKYGEFLAEKAKYPEM 1066
QY 994 FTKQISTQOAIQDPSVKIKOWSARYFNNGSILGRGADYVLSQVSNKYFNVASDITFLPS 1053
Db 1067 FKYNMISTGKPIDDSVKLKQWKAKEYFNGTVNLERGCVGLSDEATGKYFTVTKEGFIPL 1126
QY 1054 SLGK-VWESGIRYDGKGYLYNSSATGQDQKASFIITEAGNLHYFGKDGVMWTAQ-TING 1111

Db 1127 QLTGKEKIVTGFSSDGKGIY--FGTSGTQAKSAFVTFNGTYYFDARGHVTNSEVSPNG 1185
QY 1112 AN-YFPLENGTALRNTIYTDAGNSHYANDGKRYENGYYQF-----GND-----WR 1157
Db 1186 KDVYRFLPNGIMLSNAFYIDANGNTYLYNSKQMYKGGYTKFDFVSETDKGKESKVVKFR 1245
QY 1158 YF-KGNMAVGLTTVDGNVOYFDKGVQAKDKLIIVTRDGKRVYFDHNGNAATNTFIADK 1216
Db 1246 YFTNEGVMAGVTVIDGFTQYFEGDFQAKDK-LVTFKGKTYTFDAHTGNGIKDTW-RNI 1303
QY 1217 TGHWTYLGKGVAVTGAOTVKGKLYFEANGQOVKGDFTVSDGKLYFYDVSDDMWTDIT 1276
Db 1304 NGKWTYFDANGVAATAQAVINGQKLYFNEGSGQVGGVKNADGTSYKKEGFGELVTNE 1363
QY 1277 FIEDKAGNWFYLGKDGAAVTGAOTIRGQKLYFKANGQQVKGDIVKGTGDKIRYYDAKSGE 1336
Db 1364 PFTTDGNVWYAGANKTGTGAOVINGQHLFYFNADGSQVKGGVKNADGTSYKYNASTGE 1423
QY 1337 QVFNKTVKAAADKTYTVIGNDGVAVDPSVVKGTFFKDDASGALRFYNLKGQLVTSGHWYETA 1396
Db 1424 RLTFNE-----PFTTG 1433
QY 1397 NHDWVI-QSGKALTGEQTINGOHLFYFKEDGHQVKGQLVGTGDKVRYYDANSDDQAFNK 1455
Db 1434 DNNWYIIGANGKSVTGEVKIGDITVFFAKDGKQVKGQTVSAGNGRISYYTSGSGKRAVST 1493
QY 1456 SVTVANGKTY-YFGNDGTA 1472
Db 1494 WIEIQPVVYVYFDKNGLA 1511

RESULT 15

US-10-797-821-40
; Sequence 40, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-797-821-40

Query Match 39.8%; Score 3325.5; DB 5; Length 1518;
Best Local Similarity 45.0%; Pred. No. 2.3e-173;
Matches 710; Conservative 239; Mismatches 456; Indels 173; Gaps 33;
QY 1 MEKNVRFKHKVKKRVTLSVASATMLASALG-----ASVASAD-----TDTASDD----- 46
Db 1 MENKIHYKLHKVKKOWVTIAVASVA-LATVGLGLSVTTSVSADETQDKTKVTQSNSTGA 59
QY 47 ----SNOAVVTGDTNNQATD-----OTSIAATAT-----SEOSASTDAATDOASAA 90

Db 60 SLVTSPEATKEADKRTNTEADVLTPAKETNAVETATTTNTQATAEAAATATTADVAVA 119
Qy 91 EOTQGTASTDTAAQTITNANE-AKWPTENENOGFTDEMLAEKQNVATAESDISPSDLA 149
Db 120 VPKAEAVTTTDAVAVTEKAEQPAVKAENVN-----TEVKAPEALKDSEVEAL 171
Qy 150 KMSNVQVDKYYKYDODGNVKNFVAVSGDKIYYFDETGAYKXDTSKVDADKSSVASQN 209
Db 172 SLKNIKNIDKYYVNVNEDSGHKNFALTNGQLLYFGKOGALTSSSTYSPTGTTNVDG 231
Qy 210 ATTFAANRAYSTSAKNFEAVDNYLTADSWYRPSILKDGKWTSTESKODFRLLMAWPP 269
Db 232 ---FSINNRAYDSEASEFELIDGLTADSWYRPSIIKDGVWQASAEADFRLLMAWPP 288
Qy 270 DTETKRNYNMKNVGVGIDKTYTAETSQADLTAAAEVLQARIKQITSENNTKWLREAS 329
Db 289 NVDTOVNYLNMKSVNLDKAYSSDQKQETLVAAKDIQIKIEQKQAEKSTQWLBETIS 348
Qy 330 AFVKTOPQWNGESEKPY-----DHLQNGALLFDNQTDLTPDTQSNRYLLNRTPNTQTS 384
Db 349 AFVKTOPQWNGESEN-YSKGGGDHLLQGGALLYVND-SRTPWANSYRRLNRTPNTQGT 406
Qy 385 LDRSFTYNDP--LGGYDPLANDVNSNPVQAEOLNWLHLLNPGSIYANDADANFD 442
Db 407 IDKSLDEOSDPNMGDFLLANDVDLSNPVQAEOLNQHLYLMNMGSIYMGDKDANFD 466
Qy 443 SIRDAVDNVDALLQISSDYLKAAYGIDKNNKANNHVSIVAEWSDNDTPYLHDDGDNL 502
Db 467 GIRDAVDNVDADMLQLYTNYFREYGVVNVKSEANALAHISVLEAWSLNDHNDKTDGAA 526
Qy 503 MWNNDKFRSLMWSLAKPLDKGS-GLNPLIHSNLDREVDDR-----EVEV 548
Db 527 LAMENKQRLALLFLAKPIKERTPAVSPLYNNTFNTTQDEKTDWKNGSKAYNEDGTV 586
Qy 549 P-----SYSPARHDSVODIIRDIKAEINPNSGYSTQEEIEOAPK 592
Db 587 KOSTIGKYNKYGDSAGNVVFRADHNNVDIIAEIIEKKEINPKSDGFTITDAEMKQAFE 646
Qy 593 IYNEDLKTDKKTTHYNPLSYTLNLTNGSIPRVVYGMFTDDGGYMANKTWNYDAIES 652
Db 647 IYNKMLSSDKKYLNNIIPAAVAMLQNMETITRVVYGLYTDGHHMETKSPYYITV 706
Qy 653 LKARKKYYGSGOAMQNYQI-----GNGEILTSTRYKGGALKQSD-KGDAT 697
Db 707 LMKSRIKYVSGGOAQRSYWLPDGMKNSDVELYRTNEVTVSTRYKGDIMTANDTEGSKY 766
Qy 698 TTSVGVMGNQPNFSLDGKV-VALMGAHANQBYRALMVSTKDGVAITYATDADASKA 756
Db 767 SRTSGOVLVANNPKLNLDSQAKLANVEMGKIHANQYRALIVGTADGIKNFTSDADAI 826
Qy 757 GLVKRTDENGILYFLNDDLKGVANPOVSGFLOVWVPVGAADDQDIRVAASDTASTDGK-S 815
Db 827 GYVKETDSNGVLTFGANDIKGYETDMSGFVAVWVPVGASDNDQIRVAPSTAKKEGELT 886
Qy 816 LHQDAMDSRVNMEFGFSNPSF--ATKEBEYTVNVIANNVDKFSVNGITDFEMAPQYVSS 873
Db 887 LKATEAYDSQLIYEGFSNFOTIPDGSDPVYNNRKAENVNDFKSGVTSFEMAPQYVSA 946
Qy 874 TDQQLDSVIQNGYAFTRDYDLMGSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMY 933
Db 947 DDGTFDLSVIQNGYAFADRYDALMSKNKNGSKEDLDALKALHAKGIAQIAIADWVPDQY 1006
Qy 934 TPKQGVTVTTRDKFGPIAGSOINHSLYVTDTKSSGDDYQAKYGGAFDELKEKYP 993
Db 1007 QLPKGEVTAITDAGRKIAIDAIIDHSLYVANSKSSGKYQAKYGGEFELAKKAKYPEM 1066
Qy 994 FTKQISTQQAIDPSVKIKQWSAKYFNGSNIILGRGADYVLSQVSNKYFNVASDITLFLPS 1053
Db 1067 FKVNIMSTGKPIDDSVLKQWKAKEYFNGTNVLERGVYVLSDBATCKYFTVTKGNFPL 1126
Qy 1054 SLLGK-VVESGIRYKGGKYYINSSATGQVKASFITEAGNLVYFGKGGYMTGAQ-TING 1111
Db 1127 QLTGKEKVTGFSDDGKITTY-FGTSGTQAKSAFVTFNGNTYYFDARGHMTNSBSPNG 1185

Qy 1112 AN-YFLENGTALRNTIYTDAQGNSHYANDGKRYENGYOOF-----GND-----WR 1157
Db 1186 KDYRFLPFGIMLSNFAFYDANGNTYLYNSKGQMYKGGYTKFDFVSETDDKDKESKVVKFR 1245
Qy 1158 YF-KDGMVAVGLTVDGNVQYFDKDGVOAKDKIIVTRDGRKRYRFDQHNGNAATNTFIADK 1216
Db 1246 YFTNEGVMAGVTVIDGFTQYFGEDGFOAKDK-LVTFKGTYYFDAHTGNGIKDTW-RNI 1303
Qy 1217 TGHWYLLKDGVAVTGAQTVGKOKLYFEANGQOVKGDFTVSDGKLYFYDVSDDMWDT 1276
Db 1304 NGKYYFDANGVAATGAQVINGOKLYFNEGDSQVKGWKNADGTVSKYKEGFEELVNE 1363
Qy 1277 FIEDKAGNWFYLGKDCGAAVTGAQTVIRGOKLYFKANGQOVKGDIVKGTGDKIRYVDAKSGE 1336
Db 1364 FFTDGNWYVYAGANKTVITGAQVINGQHLVFNADSGOVKGGVVKRADGTVSKYNASTGE 1423
Qy 1337 QVENKTVKAADGKTYVYVIGNDGVAVDPVVKGTFFDASGALRFYNLKGQVLTGSGWYETA 1396
Db 1424 RLTVNE-----PFTTG 1433
Qy 1397 NHDWYI-OSGKALTGEQTINGQHLVFKEDGHQVKGQVLTGTDGKRYRYDANSQDQAPNK 1455
Db 1434 DNNWYIYGANGKSVTGEVKGIDDTYFFAKDGKQVKGQTVSAGNGRISYYYGDSGKRAVST 1493
Qy 1456 SVTVNGKTY-YFGNDGTA 1472
Db 1494 WIEIQPGVTVYFDKNGLA 1511

Search completed: February 11, 2006, 20:57:01

Job time :151.315 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 9.04803 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-37
Perfect score: 8349
Sequence: 1 MEKNVRFQHKVKRWVLS.....VGYQYFFGNGARIYRGWN 1590

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New:*
1: /cgn2_6/prodata1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/prodata1/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/prodata1/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/prodata1/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/prodata1/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/prodata1/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/prodata1/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/prodata1/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	513	6.1	2710	US-11-051-453-41	Sequence 41, Appl
2	443.5	5.3	726	US-11-052-554A-217	Sequence 217, Appl
3	381	4.6	2367	US-11-051-453-42	Sequence 42, Appl
4	285.5	3.4	2314	US-11-013-759-11	Sequence 11, Appl
5	284.5	3.4	2053	US-11-013-759-9	Sequence 9, Appl
6	279.5	3.3	1992	US-11-013-759-13	Sequence 13, Appl
7	279.5	3.3	2047	US-11-013-759-4	Sequence 4, Appl
8	279.5	3.3	2047	US-11-013-759-7	Sequence 7, Appl
9	279.5	3.3	2047	US-11-013-759-13	Sequence 13, Appl
10	265	3.2	1588	US-11-052-554A-280	Sequence 280, Appl
11	258.5	3.1	5024	US-10-793-626-2364	Sequence 2964, Appl
12	257	3.1	396	US-11-022-562-228	Sequence 228, Appl
13	254.5	3.0	658	US-10-873-528-17	Sequence 17, Appl
14	254.5	3.0	677	US-10-873-528-155	Sequence 155, Appl
15	249.5	3.0	1155	US-10-793-626-1780	Sequence 1780, Appl
16	248	3.0	619	US-11-052-554A-229	Sequence 229, Appl
17	242.5	2.9	3194	US-11-052-554A-90	Sequence 90, Appl
18	241.5	2.9	5291	US-11-052-554A-281	Sequence 281, Appl
19	236.5	2.8	690	US-11-052-554A-232	Sequence 232, Appl
20	234	2.8	2902	US-11-052-554A-91	Sequence 91, Appl
21	228	2.7	1448	US-10-485-517-212	Sequence 212, Appl
22	226	2.7	701	US-11-052-554A-231	Sequence 231, Appl
23	225	2.7	744	US-10-873-528-184	Sequence 184, Appl
24	223.5	2.7	1647	US-11-052-554A-260	Sequence 260, Appl
25	214	2.6	338	US-11-052-554A-228	Sequence 228, Appl

26	212	2.5	2399	7	US-11-052-554A-92	Sequence 92, Appl
27	211.5	2.5	1461	7	US-11-052-554A-283	Sequence 283, Appl
28	209	2.5	483	7	US-11-113-775A-1	Sequence 1, Appl
29	209	2.5	483	7	US-11-113-799-1	Sequence 1, Appl
30	209	2.5	1571	7	US-11-052-554A-2	Sequence 2, Appl
31	207.5	2.5	501	6	US-10-630-203-25	Sequence 25, Appl
32	207	2.5	1562	7	US-11-052-554A-211	Sequence 211, Appl
33	206	2.5	483	7	US-11-113-775A-2	Sequence 2, Appl
34	206	2.5	483	7	US-11-113-799-2	Sequence 2, Appl
35	203	2.4	980	7	US-11-052-554A-17	Sequence 17, Appl
36	202	2.4	1767	7	US-11-052-554A-372	Sequence 372, Appl
37	201	2.4	485	6	US-10-630-203-12	Sequence 12, Appl
38	201	2.4	485	7	US-11-113-775A-3	Sequence 3, Appl
39	201	2.4	485	7	US-11-113-799-3	Sequence 3, Appl
40	199	2.4	2340	7	US-11-052-554A-171	Sequence 171, Appl
41	198	2.4	1345	7	US-11-052-554A-282	Sequence 282, Appl
42	197	2.4	1263	6	US-10-485-517-127	Sequence 127, Appl
43	196	2.3	1107	6	US-10-485-517-145	Sequence 145, Appl
44	195	2.3	501	6	US-10-630-203-27	Sequence 27, Appl
45	195	2.3	501	6	US-10-630-203-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1

US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: ZHANG, JR., WILLIAM D.
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; PRIOR FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match	6.1%;	Score 513;	DB 7;	Length 2710;
Best Local Similarity	21.2%;	Pred. No. 6.9e-20;		
Matches 364;	Conservative 242;	Mismatches 619;	Indels 496;	Gaps 92;
QY	119	ENENQFTDE--MLAEA-KNVATAESDSTPSDLAKMSN--VKQVD-----GKYY----	162	
Db	1211	ETENLDFSKIMLMNAPSRSRVFWETGAVPG-LRSLNDGTRLDSDIRLYPGKFTWRFY	1269	
QY	163	YYD-----QDGNVK-----KNFAV-----SVGDKI-YYFDGTG-----	189	
Db	1270	AFDYVAITTLKPYVEDTNIKIKLDKXDRNFIMPTTNEIRNKLVSFSGAGTYSLLS	1329	
QY	190	AYKOTSKVDADKSSAVSQNATIPFANNRAYTSKAFVADNYLTADSWYR-----PKSI	245	
Db	1330	SYPISTINILNSL-----DDLWIFNDREVISIENGTIKKGKLIKDVLSKIDINKNL	1383	

Db 198 SDGNKHYYDRDSEMTNRNFVNDQ--GNWYLLNDGVVPVTSISITVNGOSLYFNSDGSQVKG 256
QY 1318 DIVKGTGKIRYYDAKSGEQVFNKTKAADGKTYVIIGNDG--VAVDP-SYVKGQTFKDA- 1373
Db 257 NFVE-EDGSLRYDKNSGD--LLRKTSTINGVNYQFPDNDGNARAIDKIEVVKTSLVVDSY 314
QY 1374 -----SGALRFVNLK-----GQVLTSGHWYETANHDWVYIQSGKALTGSGQTNGOH 1419
Db 315 EFGPSVKIILEFNHKKVTPAVVHAGAMVTTAGVQRKILNSVYNSASG-----H 362
QY 1420 LYFKEDGHQVKGOLVGTGDKVRYDANSQDAFNKSVTVNGKTYTFGNDGTAQTAGNPK 1479
Db 363 VTFDSSHVYTLSDIP-----YDPN---DSSRNASPFIFDSAAFRNNMWNVSYTVKVDN 413
QY 1480 GQTFKDGSD 1488
Db 414 LQVQADGSN 422

RESULT 3
US-11-051-453-42
; Sequence 42, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-PEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; FILE REFERENCE: MJI-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; PRIORITY FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 42
; LENGTH: 2367
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-42

Query Match 4.6%; Score 381; DB 7; Length 2367;
Best Local Similarity 21.1%; Pred. No. 8.3e-13;
Matches 292; Conservative 203; Mismatches 474; Indels 412; Gaps 82;

QY 356 LFDNOTDLPTDQTS-----NVRLLNRPTNTQSLDSRF-----TVPNDPLGGY- 400
Db 1230 VFAMEGTGTPGLSLENDGTCLLDRIRDNYEGEFYRYFAFTADALITLKR-----YE 1284
QY 401 DFLANDVDSN-----PVQA-----EQLNWLHYLLNFGSIYA-----NDADANFD 442
Db 1285 DTNIRINDSNRSFVPIITTEYIREKLSYSPY--GSGGTVALPLSQYMMGINIELSES 1342
QY 443 SIRVDVNDVNDALLOISDYLK-----AAYGIDKNNKNNHNSVIVEAWSNDTP 493
Db 1343 DWMIIDVNDVNRD-VTIESDKTKKGLIEILSTLSIEENKIILNSH----- 1388
QY 494 YLHDDGNLMMNDKFRLLSMLAKPLDKRSGNPLIHNSLVREVDDREVETVPSYSF 553
Db 1389 EINFSGE--VNSNGF-VSLTFSILE-----GINAIEVDLSK-----SYKL 1428
QY 554 ARAHDSVODIIRDIKAEINPNSFGYSFTQEBIEQAPKIYNEDLKKTDKKYTHNVPLS 613

Db 1429 -----LISGELKILMLNSNHIQQKID--YIGFNSBELQK-----NIPYS 1464
QY 614 YTELLLTNKG- - - - -IPRVYGDMTFTDDCOYMAN-----KTVN 646
Db 1465 F--VDSSEKENGFIINGSTKEGLFVSELDPVVLISKVYMDDSKPSFGYYSNNLKDKVVIIT 1521
QY 647 YDAIESL-----LKARMKYVSGGQAMQNY-----QIGNGEILTSVRYGKALKQS 691
Db 1522 KDNVNILTYGKLDKIK-ISLSLTQDEKTIKLSNHLDESQVAEILKFWN-RKGSTNWS 1579
QY 692 DKGDATRRISGVGVWGN--OPN--PSLDGKVVALNMGAAHQAQYERALAMVSTKDGVATY 747
Db 1580 DSLMSFLESNWKISIFVNFLOSNIKFILDFANFI-----ISGTTISGQFEP 1624
QY 748 ATDADAS-KAGLVK-RTDENGILYFLNDDLKGVANQVSGFLQVWVPVGAADQDQIRVAA 805
Db 1625 ICDENNNIOPYFIKFWTLENTYLYVGNRQNMIVEPYD-----LDDSGDI---S 1671
QY 806 SDTASTDGKSLHODAAAMDSRV-----MFEFGSNFQSPATKEEBEYTNVI--ANNVDK 855
Db 1672 STVINFSQKLY--GIDSCWKNKVISPNIIYTBINITPVYETNNTYPEVIVLDANYINE 1728
QY 856 FVSWGITDFEMAPQYVSSTDGQ-----FLDSVIQN--GVAPT 890
Db 1729 KINVNINDLSI--RYVMSNDGNDFILMSTSEENKVSQVKIRFVNVFKDKTLANKLSFNFS 1786
QY 891 DRYDLGMSKANKYGTADQLVKAIKALHAKGL-----KVMADWVPDQMYTFPKQEVVTVRT 946
Db 1787 DKQDVPVSEI-----ILSFTPSYEDGLIGYDLGLVSLYNEKFY-----I 1826
QY 947 DKFGKPIAG-SQINHSLY-----VTDTKSSGDDYQAKY-----GGAFDELKBEKY 990
Db 1827 NNFGMMVSGLIYINDSLYFKPVPNNLIITGFVTGDD--KYVFNPIINGCA----- 1874
QY 991 PELFTYKQISTGAIDPSVKIKOWSAKYFNNGSILGRGADYVLSQVSNKYFNVASTDLP 1050
Db 1875 -----ASIGETI---IDDKNY---YFNQSGVLQTG--VFSTEDGKFYFAPANT--- 1914
QY 1051 LPSLLGKVVESGIRYDGGKLYNSSATGDQVKASFIETAGNLYYFGKQGYMTGA--Q 1107
Db 1915 LDENLEGAID---FTGKLI---DE-----NIIYF-EDNY--RGAVEWK 1950
QY 1108 TING-ANYFFLENGTALR--NTIYTDAQNSHYIANDGKRYENGYOQFGNDMRYFKD-GN 1163
Db 1951 ELDDGEMHYFSPETGAKFGLNLI-----GDDKYFNSDGMQKGFVSINDKNHYFDDSGV 2005
QY 1164 MAVGLTTVDGNVOYQPKQVQAKDKIIVTRDQKRVYFDQHNHNAATNFIADKTGHYYL 1223
Db 2006 MKVGYTEIDGKHIFYAENG-EMQIGVFNTEDG-FKYFAHNED-----LGNEEG----- 2052
QY 1224 GKDGVAVTGAQTVGKQKLYFEANGQOVKGFVTSDEGKLYFYVDVSDGMWTD-TFTEDKA 1282
Db 2053 --EBSISYGLNPNKNIYYFDDSFATAVG-WKLEDGSKYIFDEDTAEAYIGUSLIND-- 2107
QY 1283 GNVFYLKDGAAVTCATQITRQKLYFKANGQQVKGDIIVKGTGDKIRYDAKSGEQVFNKT 1342
Db 2108 -QOYFNDGDIQVGFVTVINDKVFFSDSGIIEBSG--VQNIIDNYYFIDNGIVQI--GV 2162
QY 1343 VKAAGKTVVIG-----NDGVAVDPSVVKGTQFKDASGALR-----FYNLKGQVLTGSGW-Y 1393
Db 2163 FQTSQGYKYFAPANTVNDN-----IYQAV-EYSGLVRVGDDVYFYFETTYIETGIWY 2214
QY 1394 ETANHDWVI---QSGKALTGQTINGOHLKPKEDGHQVKGQVLTG--TDGKRVYYDANS 1448
Db 2215 DMENESDKYFYFETPKACKACKINLIDDIKYFDE-----KGIMRTGLIIFENNNYFNEN 2269
QY 1449 GQOAFNKSVTYNGKTYFVGNDCGTAQTA--GNPKGTQFKDGSDIRFYSMGQVLTGSGWYE 1506
Db 2270 GIIQFG-YINIEDKMFYFGEDGMVQMGVNTDGPVKYFAHQNTLDENFEGESINYTGW-- 2326
QY 1507 NAQGOWLYYKNGKVTLTGQTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDENSGSMITN 1566
Db 2327 -----LGLD---EKRYFTDEYIAATGSVI--IDGBEYFDPDQALVIS 2366

QY 1567 Q 1567
Db 2367 E 2367

RESULT 4
US-11-013-759-11
; Sequence 11, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-11

Query Match 3.4%; Score 285.5; DB 7; Length 2314;
Best Local Similarity 19.0%; Pred. No. 1.2e-07;
Matches 340; Conservative 203; Mismatches 673; Indels 571; Gaps 77;

QY 29 SALGASVASADTTDASD-----DSNQAVVTGDTTNNQATDQTSIAATATSEQASSTD 81
Db 792 NATTAKVYDETNQTSKVTVVNVDEKTIETLGDNGKTKIGVKTTLTTTNNANGKATNF 851
QY 82 AATDQ-----ASAEQOTGTTASTDQAQT-----TT 108
Db 852 STTDNDALVNAKDIAENLNLTAKEIHTTGTADTALQTFVKKDGATDDETTIVGKDGTO 911
QY 109 NANEAKWVPTEENOGFTDEMLEAKNVATAESDSIPSLAKSNVQKVDGKYVYDQDG 168
Db 912 NGKTVNTLKLKGN-GLT-----VATNKDGTVTFGINTOSGLKAGDST--TLNKDG 959
QY 169 ----NVKKNFAVSVG-DKIYFDETGAYKDTSKVDADKSSAVSQNATIPAAANNRAYSTS 223
Db 960 LSIKNPASNEQIQVAGGVKF-----AKVDKGNSTGIDGTSRI--TKDQIGFTGA 1008
QY 224 AKNFEADVNTLTADSWYRPSKILKDKGTWESGKDDFRPLLMAWPDTETKRNVNY-MN 282
Db 1009 NGSLDTPKPLTKDK-----LKVG-----EVEITNTGINAGGK 1041
QY 283 KVGIGDKTYTAETSQADLTAABEL-VQARIEQKITSENTNKTWLRREALSAFVKTPQOWN-- 339
Db 1042 KITNIOSGDITONSNDNAVGTGRVYDLKTELESKINSA--AKTAQNSLHFEVSADQGNHF 1099
QY 340 --GESEKPYDHLQNGALLFDNOTDLPDQSNVRLNRPPTNQTGSLDS-RFTYNPNDP 396
Db 1100 TVSNPYSYDTSKTSVDITFAGENGIT--TKVNGVV-RVIGDQTKGLTTPKLTGVNNG 1156
QY 397 LGGYDFLLANDVNSNPVQAEQLNMLHYLLNFGSIY-----ANAD----- 438
Db 1157 KG-----IVIDSQKQNTITGLS--NTLANVTNDGAGHALSQGLANDTDKTRAASIGDVLN 1210
QY 439 -----ANFDSI-----RVDADVNDADLLQISSDY-----LKAAYGIDKNNKNNHVS 482
Db 1211 AGFNLGNGEAVDFVSTYDTVDVFDGNATTAKTVDYDTSKTSKVYDVNVNDKNT----- 1264
QY 483 IVEAWSQ-----NDTPYLHDGDGLNLMNMDNKFRLSMLWSLAKPLDKRS 525
Db 1265 -IEVTSKGLGVKVTTLTKTSANGNATKFSAADGD-----ALVKASDIAT 1308

QY 526 GLNPLHNSLVDRVDREVTVPSPYSPARAHDSFVQDIIRDIILKAEINPNSFGYSFTQE 585
Db 1309 HLNTLA-----GDIQTAKGASQASSASYV-----DADGNKIYD----- 1343
QY 586 EIEQAFKIYNEDLKTKDKKYTHYNVPLSYTLTLTNKGSIPRVVYGMFTDDGQYMANKTV 645
Db 1344 -----STDKKYYQVN-----DKQVDKNKEV 1364
QY 646 NYDAIESLLKARKMYVSGQAMONY-----QINGEILTSVRYGKGALQKSDGDATT 698
Db 1365 AKDK-----LVAQAQTPDGTLAQMANVKSVINKEQVANDANKQGINEDNAFIGLENAAKDT 1420
QY 699 RTSGVGVVMGN-----QPNFSLDGKVVALLNMGAAHQAHOEYRALMVSIKDGVTATDA 751
Db 1421 KTKNAAVTVGDLNANAQAQTELTTFAGDTGTTAKKLGE-----TLTIKGGQTTNKL 1469
QY 752 DASKAGLVKRTDENGLYFLNDDL-----KGVANPQVSGFLQVWVPVGA 796
Db 1470 TNNIGVVAQTD--GETVKLAKDLTNLNSVNAGGTRIDEKGISFVDANGQAKANTPVL 1527
QY 797 DDQDI-----RVAASDTASTDGKSLHQAAMDSRVMEFGFSNFQSFATKEBEYTNVIVANNV 853
Db 1528 NGLDLGGKRI SNIGAAVDDNDV-----NFKQFNEVAKTVNNLNQNS 1571
QY 854 DKFVSWGITDPFEMAPQYVSTGQFLDSV--IQNGYAFTRDYDLGMSKANKYGT-ADOLV 910
Db 1572 GASLPFVVTDANGK--INGTDGPKQKAIKAGADGKYHANANGVPPVDKGPITDADKL- 1628
QY 911 KAIKALHAK-----GLKMADWVPDQMYTPPKQEVVTVTR-----TDKFGKPIAG-- 955
Db 1629 -ANLAHAGKPLDAGHQVVAS-----LGNSDAITLNIKSTLPQIDPTNGANAGQA 1680
QY 956 -----SQINHSLYVTDTKSSGDYQAKYG-----GAFIDELKEKY 990
Db 1681 QSLPSLSAAQQAASVKDLNVGNFLQTNHNQVDFVKAYDTVNFVNGTGADITSVRSAD 1740
QY 991 PELFTYKQISTQAI-----DPSVKIKQMSAKYF-----NGSNILGRGA----- 1029
Db 1741 G-----TMSNITVNTALAAATDDGCVLIIKAKDGKPYKADLMPNGSLKAGKSASDAKTPTGL 1797
QY 1030 -----DVLSDQVSNKYF-----NVASDTLFLPSSLLGK-----VVE 1061
Db 1798 SLVNPAGKSGTGDAVALNNLSKAVPKSKDGTITTTVSSDGI-----SIQKONSSITLSK 1853
QY 1062 SGIRYDGK--GYIYNSSATGDQVKASFITEAGNLYYFGK-----DGYVMTGAQTINGAN 1113
Db 1854 DGLNVGKGVISVNGKTKDIDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKDPN 1913
QY 1114 YFFLENGTALRNTIYTDAGNSHYANDGKRYENGYYQFGNDWRYFKDGNMAGVGLTIV-- 1171
Db 1914 SGSSNRVTIKAGTVLGGKN-----NDTEKLATGVQVGVVD-----KGNANGDLNVMV 1964
QY 1172 -----DGNVOYFDKDGVOAKDKIIVTRDGVKRYPDQHNAGNAATNTFIA 1214
Db 1965 KTKQDSKALLATYNAAGQTYLNTNPAEALDRI---NEQGIRFFHVDNGQEP----- 2016
QY 1215 DKTGHYYLKGDAVAVTGAQTVGKQKLYFEANGQQVKGFVTSDEGLKYFYVDVSGDMWT 1274
Db 2017 -----VWGRNGI---DSSASGKHSV---AIGFOAKAD-----GEAAV 2048
QY 1275 DTFIEDKAGWVFLGKDAVAVTGAQTVGKQKLYFKANGQQVKGDIVKGTGDKIRYIDAKS 1334
Db 2049 AIGRQTOAGNQSIAGIDNAQATGDOSI-----AIG---TGNVTG-----KHS 2088
QY 1335 GEQVFNKTKAADGKTYVIGNDGVAVDPSVVKQTFPKDASGALRFYNLKGQLVTGSGWYE 1394
Db 2089 GAIGDPTVKA--DNSYSYVGNNOQFIDAT-----QTDVFGVNNITTESNSVALGSNAI 2142
QY 1395 TA-NHDMVVIQSGKALTGEQTINGQHLYPKE-DGHQVKGQLVGTGDTG-KVRYIDANSQDQ 1451
Db 2143 SAGTHAGTQAKKSDGTAGTITTTAGATGTVKGFAGQAVGAVSVGASGAERRIQNVAAGEV 2202

```

Qy 1452 AFNKSVTWNGKTYFEGNDGTAQTAGNPKGQTFFKDGSDIRFYFVMEQG 1497
Db 2203 SATSTDAVNGSOLYKATOGIANATNE-----LDHRIHQNEKANAGISSAMAWASM 2253
Qy 1498 -----LVTGSGWTYENAQ-----GQWLYVNGKVKLT 1522
Db 2254 PQAVIPGRSMVTGGIATHNGQCAVAVGLSKLSDNGQWVFKINGSADT 2300

RESULT 5
US-11-013-759-9
; Sequence 9, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2053
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-9

```

QY	496	-----HDDGNLNMNDKFLRSLMLWSLAKPLDKRSGNLNPLIHNSLVDREVDD--REV	545
DB	783	NRLVIEQVPSADGNSKNI-1KGLSPLTSPASTP-----SGRNIALGNTIEEKKNSAASI	837
QY	546	ETVPSYSFARAHDESEVODIIRDI1KAE-INPNSFGYSTOEIEBQAFKI-YNEDLKKTDK	603
DB	838	DDVLNAGENLXNGKDKDFVSTYDTVDFIGDNATATATVYDEANQTSKVAYDVNVDEKTI	897
QY	604	KYTHYN-----VPLSYTLLLTNKGSIPIRVYVYGMFTDDQGYMANKTVNYDAIESLLKA	656
DB	898	ELTGDNGKQLGVKTIKLTETSTNGNA-----TTFSTDDDDHALVKASDIAGNLNTLAE	950
QY	657	RMKVYSG--GOAMQNYQINGEILITSVRYGKALKQSDKGDATYR--TSGVGVVMQGNPNF	713
DB	951	EIH7TKGTANTALFTTKK-----VDENKADDTNAITVGDGTSGKVNTL	997
QY	714	SLDGKVALNMGAAHANEYRALMVST-KDGCVATYATDADAS-KAG-----LVKR	761
DB	998	KLKK-----NGLDIKTDKGTVTFGINTOSGLKAGDSTTLNNGLSIKN	1042
QY	762	TDENGILYFLNDDLK-----GVANPQVSGFLQV-----WVPVGAADDQDIRVAASDPTA	809
DB	1043	TASNEQIQVGADGVKFAVMNVGVGAGIDGTTRITRDEIGFTGTNGSLDKSPHLXSKI	1102
QY	810	STDGKSLHQDAMDSRVWFEGFSNFQSFATKEEBEYTNVVIANNVDKPVSWGITDFEM-AP	868
DB	1103	NAGGKI-----TNIQS-----GEIAKNSHDAVTGKGIYDLKTELE	1138
QY	869	QYVGSST-----DGOPLDSVIONGYAFTDRYDLGWSK-----ANKYGTADQL	909
DB	1139	NKISSTAKTQNSLHEFSVADQNNNTVSNPYS---SYDTSKTSDDVITFAGENGITTKV	1195
QY	910	VKAIKAL-----HAKGLKVMADWVPDQMTYFPKQEVVTVTRTDKFGKPIAGSQIHNSLXYTD	966
DB	1196	NKGVVRVGIDOTKGL-----TTPK-----LTVGNNGKGIVINSQNGQNTITG	1238
QY	967	-----TKSGDDYQAKYGGAFDLBELKYPPELFTKKQI-----STGOAIDPSVKIKQ	1013
DB	1239	LSNTLANVTDKGSVRTTEQGNIIKDBDKTRAASIVDVL\$AGFNQNGEAVD---FVST	1295
QY	1014	WSAKYFNGSNILGRGADYVLSQVSNKYF--NVASDTLFLPSSILGKVVESGIRYDKGG--	1070
DB	1296	YDTVNFANGNTTAKVTYDTSKTSKVYDVNVDDTTIEVKDKKLG-VKTTTITSTGTGA	1354
QY	1071	--YIYNSSATCQO-VKASFI-----TEAGNLXYFGKGYMVTGAQTINGANYPFLENGTA	1122
DB	1355	NKPALSNAQTGDALVKASDIVAHLNTLSGDI-----QTAKGAS-----QA	1394
QY	1123	LRNTIYTDAGNSHYANDGKRVYNGQOQFNDWRVPKQGNMVAGLTTVDGNVOYFDKQ	1182
DB	1395	NNSAGYVDADGNKVIYDST---DNKYQAKN-----DGTV---DKTK	1430
QY	1183	VOAKDKLIV---TRDGKRVYFDQHGNGNAATNTFIADKTGHVYLGKGVAVTG---AQTV	1236
DB	1431	EVAOKDLVAQAQTPDGTLA---QNVKSVKIN\$QVNDANKKQGINEDNAFV\$GLEKAA\$D	1487
QY	1237	GKOKLYPEANGQOVKGDFVTSDBGKLYFYDVEDSGDMWTDTFIEDBKAGNVPYLGKGA\$VT	1296
DB	1488	NKTK-----NAAVTVGDLNAVAQTLTF---AGDTGTTA---KKLG-----EFLTIK	1528
QY	1297	GAQTI\$RQKLYFXKANGQOVKGDIVKGTGDKIRYDAK\$G\$QVFNKTVK\$ADGKTYV---IG	1354
DB	1529	GGQTDTN-----KLTDNNI--GVVAGTGD-----FTVKLAKDLTNLNSVN	1566
QY	1355	NDGVAVDPSVVKGTQPKDASGALRF-----YNLKGQLV---G\$GWYETANHDWYI	1403
DB	1567	AGG7KIDE---KGISFYDANGQAKANTPVL\$ANGGLDGGKVI\$NVGKGT\$KOTDAANVQOL	1623
QY	1404	Q\$GKAL7GEQTINGOHL\$YFKEDGHQVKQGLVGTGDKVRY\$DANS\$GDOAFN\$K\$V---T	1458
DB	1624	NEVRNLLGLGNDNA-----DGNQ-----VNTADIK---KDPNSGSS--NRTV\$KAGTV	1667
QY	1459	VNGKTYFYFGNDGT\$AQT-----GNPKG-----QTFKQ\$GSDIRFY\$MEGOLVT	1500

Db 1668 LGKK-----GNNDTEKLATGQVGVVDKGNANGDLNWNVVKTKQOG-----SKKALLAT 1717
QY 1501 GSGWYENAOQMLYVRNGKVLTLGLQTVGSQVYFDENGIGQAKGKAV-RTSDGKIRYFDEN 1559
Db 1718 -----YNAAGQTYVNN-----PAAEDRINEQGIREFHVN 1749
QY 1560 SSGMITNQMWFYVQO 1574
Db 1750 DG-----NOBPFVQGR 1760
RESULT 6
US-11-013-759-3
; Sequence 3, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentJn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-3
Query Match 3.3%; Score 279.5; DB 7; Length 1992;
Best Local Similarity 19.5%; Pred. No. 2.1e-07;
Matches 354; Conservative 232; Mismatches 648; Indels 579; Gaps 91;
QY 23 SATMLASALGA-----SVASADTDTSDDSNQAVVTGD-----OTTNNOATD 64
Db 205 SSSGLAIALGAGTGAQLOQSIALGQGSVVTQSDNNSRPAYPTNQALDPKFAQTNNTKAG 264
QY 65 QTSIAATATSEQASATDAATDQASA-----ABEQGTGTASTDTAAQTNNANE 112
Db 265 PLSIGNSIKRKLIINVAGVNTKDAVNVAQLEAVVWKAKERRITFGQDDNSTDVKIGLDN 324
QY 113 AKWVPTENEOGTDEMLAEAKNVATAESDIPSDLAK-MSNVKQYDGKYYVYDQGNVK 171
Db 325 TLTIKGAETNALTDNNIGVVK-----ADNSGLKVKLAKTLNNLTENV-----TTTLN 373
QY 172 KNFAVSVGDKIYVFDGTGAVKDTSKVDADKSSSAVSONATIPAAANNRAYSTSAKNFEAVD 231
Db 374 ATTTVKVGS-----SSTTAELLSLSLFTQNTQSGSTKTVGVNGVKFTNNAEITTAAG 430
QY 232 NYLTADSWYRPSILKDGKWTBESGKDDFRPLLMAWPPDETETKRNVTNNYNNKVVG-----286
Db 431 T-----TRITRKIGFARDGVD-----EKQAPYLDKQLKGVGSAIT 468
QY 287 IDKTYTAETSQADLTAAAEVLQARIEQKITSENNTKWLREASISAFVKTQPOWNGESEKPY 346
Db 469 IDNGIDAGNKKISNLA-----KGSSANDAVTIEQLKAA-----KP- 503
QY 347 DDHLQNGALLFDNQDILTPTDQS-----NYRL-LNKTPTNQTGSLDSRFTYNNPNDPLGG 399
Db 504 --TLNAGAGISVPTTSEISVDKAGSGNTAPTNYINGVKTTELSNGTSD-KFSVKGS-----555
QY 400 YDFLLANDVDSNPVVOAEOQLNHLNFGSTYAND-----ADANFDSIRVDAVDNVDAD 455
Db 556 -----GTNNSLVTAEHL-----ASYLNEVNRNTADSAALQSFTVKEEDDDAN 596
QY 456 LLIQSSDYLKAA-----YGIDK-----NN-----474

Db 597 AITVAKDITKRGAGVSIILKLGKNGLTVAATKKDGTVTFGLSQDSGLTICKSTLNDGLTV 656
QY 475 KNANHHVSI-----VEAWSNDPTPYLHDDGD 500
Db 657 KDTNEQIQVGANGIKFTNVNGSNPGTGIANTARITRDKIGFAGSDGAVDTNKPYL--DQD 714
QY 501 NLMMNDKFERLSMLWSLAKPLDKRSGIAPLIHNSLVDRVDDREVE---TVPSYSFARAH 557
Db 715 KLOVGNVKITNTGINAGGKAI---TGLSPTL-PSIADQ---SSRNELGNTIQDKD--KSN 766
QY 558 DSEVOIIRDIKABINPNSFCYSTOBEIEQA-----FKYNEDLKKTDKKYTHYVNP 611
Db 767 AASINDILNTGFNLKNNNPIDFVSTYDIVDFPANGNATTATVTHDTANKTSKVVDVNDV 826
QY 612 LSYTLLLTNKGSIPIRVYVYDMFTDDQYWMANKTVNYDAIESLLKARMKVYSGQGMQNYO 671
Db 827 -DTTIHLTG-----TDDNKKLGKVTTKLN-----KTSANGNTATNFN 862
QY 672 IGNG--EILTSVRYGKGALK-----QSDKGDATTTRTSGVGVVMGNQPNFSLDGKVVALN 723
Db 863 VNSSEDEALVNAKDIAENLNTLAKBIHTTKGTADTALQFTTVKKVDENNADDAN--AIT 920
QY 724 MGAHAHNOEYRALVSTKDG VATYATDADASKAGLVKRTDENGILYF-----770
Db 921 VGQKANNQVNTLTILKEN-----GLNIKTDKNGTVTFGINTTSGLKAGKS 966
QY 771 -LNDDLKGVANPOVSGFLQVWVP-----VGAADDQDIRVAASDT--ASTDGKS 815
Db 967 TLNDGGLSINKNPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTRITRDEIGFTGNG-S 1025
QY 816 LHQDAAMDSRVMFEGFSNFSQSPATKEEYTNV-----VTANNVDFKFSWGITDEEM-APQY 870
Db 1026 LDKSKPHLSK--DGIN-----AGGKKTITNQSGEIAQNSHDAVTVGGKIDYDLKTBLENK 1076
QY 871 VBST-----DGQFLDSVIONGVAFTDRLGMSK-----ANKYGTADQLVK 911
Db 1077 ISSTAKTAQNSLHERSVADEQGNNTVSNPYS---SYDTSKTSVITPAGENGITTKVVK 1133
QY 912 AIKAL---HAKGLKVMADWVPDQMYTFPKQEVVTVTRTDKFKGIAGSQINSHSLYVTD-- 966
Db 1134 GVVRVGIDQTKGL-----TTPK-----LTGVNNGKGIVIDSQNGQNTITGLS 1176
QY 967 -----TKSGDDYQAKYGGAFDELKPKYPLFTKKQI-----STGQAIDSVKIKQWS 1015
Db 1177 NTLANVTNDKGSVRTTEQGNIIKDEDKTRAASIVDVLSAGFNQGNGEAVD---FVSTYD 1233
QY 1016 AKYFNGSNILGRGADYVLSDOVSNKYF--NVASDTLFLPSSLGKVGESGIRYDGKG--- 1070
Db 1234 TVNFADGNATTAKVYDDTSKTSKVVDVNVDDTIEVKDKKLK-VKTTTLTSTGTGANK 1292
QY 1071 TYNSSATGQ--VKASFI-----TEAGNLVYFGDGYMVTGAQTINGANYFFLENTALR 1124
Db 1293 FALSNOATGDLVKASDI VAHLNTLSGDI-----QTAKGAS-----QANN 1332
QY 1125 NTIYIDAQNSHYHANGKRYENGVOQFGNDHRYFKDGNMAGVLTITVDGNVQYFDKGVQ 1184
Db 1333 SAGYVDADGNKVIYDST---DNKYQAKN-----DGTV---DKTKEV 1368
QY 1185 AKDKIIV---TRDGKRVYFDQHNGNAATNTFTADTKGHWYILGKGVAVTG---AOTVKG 1238
Db 1369 AKDKLVAQATPDGTLA---QNNVKSVINKEQVNDANKQGINEDNAFVKGLEKASDNK 1425
QY 1239 OKLYFEANGQQVKGDFVTSDEGLKLYPYDVDSGDMTDTTFIEDKAGNFWLKGKGAATGA 1298
Db 1426 TK-----NAAVTGDLNAVAQTPLTF---AGDTGTTA---KKLG-----ETLTIKGG 1466
QY 1299 QTIROGLKYPKANGQQVKDGIUKGTGTGKIRYDASKGEQVFNKTKAADGKTVV--IGND 1356
Db 1467 QTDTN-----KUTDNNI--GVVAGTDG-----FTVKLAKDLTILNLSVNA 1504
QY 1357 GVAVDPSVVKGTQTFKDASGALRF-----YNLKGQLVT--GSGWYETANHWDVVIQ 1405

Db 1505 GTKIDD---KGVSFVDSGQAKANTPVLSANGLDLGGKVISNVGKGYKQTDAAANVOQLNE 1561
QY 1406 GKALTGQTINGOHLYPEKGHQVKGQLVTGTGKVRYYDANGSQAFNKSVD---TVN 1460
Db 1562 VRNLLGLGNAGNDN---ADGNQ-----VNIADIK---KDPNSGSSS-NRTVIKAGTVLG 1608
QY 1461 GKTYFFGNDCTAQT-----GNPKG-----QTFKQSDIRFPYSMEGOLVTGS 1502
Db 1609 GK-----GNNDTEKLATGGIIVGVNDKQGNANGDLSNVWVKTKQD-----SKKALLAT-- 1656
QY 1503 GWYENAGQWLYYVKNKGKVLATGLQTQVRSQVYFDENGIOAKGKAV-RTSDGKIRYFDENSG 1561
Db 1657 ---YNAAGQTNLYTNN-----PAAIDRINEQGIREFHVN DG 1690
QY 1562 SMITNQWKFYQG 1574
Db 1691 ----NOBPVVQGR 1699

RESULT 7
US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Kleih, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-13

Query Match 3.3%; Score 279.5; DB 7; Length 1992;
Best Local Similarity 19.5%; Pred. No. 2.1e-07;

Matches 354; Conservative 232; Mismatches 648; Indels 579; Gaps 91;

QY 23 SATMLASALGA-----SVASADTDASDDSNQAVVTD-----QTTNNQATD 64
Db 205 SSSLAGALGAGTGAQLOQSGIALGOGSVVVTQSDNNSRPAYTPNTQALDPKFOATNNTKAG 264
QY 65 QTSIAATATSEGSASTDAATDQASA-----ASQTQCTTASTTAAQTNNANE 112
Db 265 PLSIGNSIKRKIIINVAGVKNKTDVNVVAQLEAVVWAKERRITFGDDNDSTDKTGLDN 324
QY 113 AKWVPTENENOGFTDEMLAEAKNVATAESDPSIDLAK-MSNVKQVDGKYYVYDQGNVK 171
Db 325 TLTIKGAETNALDNNIGVKE---ADNSGLKVKLAKTLNLTENV-----TTTLN 373
QY 172 KNFAVSVGDKIYFYDETGAUKDTSKVDADKSSAVSQNATIPFAANNRAYSTSAKPEAVD 231
Db 374 ATTTVKVGS-----SSTTBALLSDSLTFTQPTNGSQSTSKTVGVNGVKFTNNAETAAIG 430
QY 232 NYLTADSWRPKSILKDKGTWTESGKDFRPLLMWWPDTETKENVYNNVKNVVG----- 286
Db 431 T-----TRITRDKITGFARDGDVD-----EKOAPYLDKKQLKQGVSAIT 468
QY 287 IDKTYTAETSQADLTAALVQARIKQITSENNTKWLREAIISAFVKTPQWNGESEKPY 346
Db 469 IINGIDAGNKKISNLA-----KGSANDAVTIEQLKAA-----KP- 503
QY 347 DDLQNGALLFNDQDLTDPDTQS-----NYRL-LNRFTPTNQSGSLDSRFTYNNPDPLGG 399

Db 504 --TLNAGAGISVTPTEISVDAKSGNVTAPTYNIGVKTTTELNSDGTSD-KFSVKS----- 555
QY 400 YDFLLANDVDSNPVQAEOLNMLHLNLLNFGSIYAND---ADANPDSIRVDAVDNVDAD 455
Db 556 -----GTNNSLVTAEHL-----ASYLNEVNRATADALQSFTVKEEDDDAN 596
QY 456 LLOISSDYLKAA-----YGIDK-----NN----- 474
Db 597 AITVAKDITTKNAGAVSILKKGKGLTVATKDGTVTFGLSQDSGLTIGKSTLNNDLGLTV 656
QY 475 KNANNHVSI-----VEAWSNDNTPYLHDDGD 500
Db 657 KDTNEIQVANGANGIKFTNVGNSPGTGIANTARITRDKIGFAGSDGAVDTNKPYL--DQD 714
QY 501 NLNMNDNKFRLSMLSLAKPLDKRSLNPLIHNLSVDREVDREVE---TVPSYSFARAH 557
Db 715 KLVQGVNKTNTGINAGGKAI---TGLSPTL-PSIADQ---SSRIELGNITIQKD--KSN 766
QY 558 DSEVDQDIIRDIIKAEINPNSFGYSFTQEEIEQA-----FKIYNEDLKTKDKYTHYNVP 611
Db 767 AASINDILNTGPNLKNNNNPIDFVSTVDIVDFANGNATTATVTHDTANKTSKVYVDVNDV 826
QY 612 LSYTLLLTNKGSIPTVYVYGDMTDDQGYMANKTVNTDAIESLLKARKVYVSGGOAMQNYQ 671
Db 827 -DTTHLTG-----TDDNKKLGKVTTKLN-----KTSANGNTATNFN 862
QY 672 IGNG--EILTSVRYGKALK-----QSDKGDATTRTSGVGVVGMGNOPNFSLDGKVVALN 723
Db 863 VNSDDEDALVNAKDIAENLNTLAKEIHTTKGTADTALQTFTTVKKVDENNADDAN--AIT 920
QY 724 MGAHAHQEYRALMVSTKDGVAITYATDADASKAGLVKRTDENGILYF----- 770
Db 921 VGQKNANNVNTLTGKGN-----GLNIKDKNGTVTFGINTTSLGKAGKS 966
QY 771 -LNDLLKGVANQVSGFLQVWP-----VGAADDQDIRVAASDT--ASTDGKS 815
Db 967 TLNDGGLSINKPTGSEIQIAGVADGVKFAKVNNGVVGAGIDGTTRITRDEIGTGTNG-S 1025
QY 816 LHQDAAMD SRVMFEGFSFQSPATKEEYTNV---VIANNVDKFSVSWGITDEM-APQY 870
Db 1026 LDKSPHLSK---DGIN-----AGKKITNQSIEIAQNSHDAVTVGGKIYDUTLEENK 1076
QY 871 VSST-----DQQLDSVIONGYAFTDRYDLGMSK-----ANKYGTADQLVK 911
Db 1077 ISSAKTAQNSLHEFSVADEQGNFTVSNPYS---SYDTSKTSDLVITPAGENGITTKVKN 1133
QY 912 AIKAL--HAKGLKVMADWVPDMYTFPKQEVVYVTRTDKFGKPIAGSQINHSLSYVD-- 966
Db 1134 GVVRVGIDQTKGL-----TTPK-----LTVGNNGKGIVIDSQNGQNTITGLS 1176
QY 967 -----TKSSGDDYQAKYGGAFDELKEKYPELFTKKQI-----STGQALDPSVKIKQWS 1015
Db 1177 NTLANVTNDKSVRTEQGNIIKDEDKTRAASIVDLSAGPNLQGNGEAVD---FVSTYD 1233
QY 1016 AKYFNGSNILGRGADYVLSDOVSNKYF--NVASDITLFLPSSLGKVVESGIRYDVGK--- 1070
Db 1234 TVNFADGNATTAKVYDDTSTKTSKVYDVNVDDTIEVXDKKLG-VKTTTLTSTGTGANK 1292
QY 1071 YTYNSATGDQ-VKASFI-----TEAGNLYYFGKDGWYVGTGAQTINGANYFFLENGTALR 1124
Db 1293 FALSNOATGDALVKASDI VAHLNLTLSGDI-----QTAGKAS-----QANN 1332
QY 1125 NTIYTDAGNSHYANDGKRYENGVOFGNDWRYFKDGNMAVGLTIVDGNVQYFDKGVQ 1184
Db 1333 SAGYVDAGNKVIYDST-----DNKYQAKN-----DGTV---DKTREV 1368
QY 1185 AKDKIIV---TRDGKRVYFDQHNNAATNTEIADTKGHWYVLGKGVAVTG---AOTVGK 1238
Db 1369 AKDKLVAOQTPDGLA---QNVKSVINKEQVNDANKQGINEDNAFVKGLEKASDNK 1425
QY 1239 QKLYFEANGQQVKGDFVTSDEGLYFYDVDSGDMWTDTFIEDKAGNWFVLGDKGAATGA 1298
Db 1426 TK-----NAAVTVGDLNVAQTPLT-----AGDTGTTA---KKLG-----ETLTKG 1466

QY 1299 QTRGOKLFPKANGQOVKGDIKVGTDGKIRYIDAKSGEQVFNKTVKAAADKTVV--IGND 1356
DB 1467 QTDN-----KLTDDNI--GWAGTDG-----FTVKLAKDLTNLSNVAG 1504
QY 1357 GVAVDPSVVKGQTFKXASGALRE-----YNLKGQVLT--GSGWYETANHDWYVYIQS 1405
DB 1505 GTKIDD---KGVSVFVSSGQAKANTPVLNANGLDLGGKVISNVCKGTCTDDAANVOQLNE 1561
QY 1406 GKALTCGEOTINGOHLYFKEDGHQVKGQLVTGTGKVRYYDANGSDQAFNKSV-----TVN 1460
DB 1562 VRNLLGLGNAGNDN---ADGNQ-----VNIADIK--KOPNSGSSS-NRTVIKAGTVLG 1608
QY 1461 GKTYFGNDGTAGTA-----GNPKG-----QTFKDGSDIRFYSMEGQLVTS 1502
DB 1609 GK-----GNNDTEKLTGCGIQGVGDKNANGDLSNVWVKTKDQ-----SKALLAT-- 1656
QY 1503 GWYENAGOWLYVYKNGKVLTLGTLVGSQRVYPBENGIOAKGKAV-RTSDGKIRYFDNSG 1561
DB 1657 ---YNAAGQTNVLTNN-----PAAIDRINEQIRFFHVNDG 1690
QY 1562 SMITNQWKFYVG 1574
DB 1691 ----NQEPVQGR 1699
RESULT 8
US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-4)
Query Match 3.3%; Score 279.5; DB 7; Length 2047;
Best Local Similarity 19.5%; Pred. No. 2.2e-07;
Matches 354; Conservative 232; Mismatches 648; Indels 579; Gaps 91;
QY 23 SATMLASALCA-----SVASADTDATSDSDSNQAVVTD-----QTTNQATD 64
DB 260 SSSLGAIAGLQAGTQALQGSALQGSVVTQSDNNRSPAYTNTQALDPKFAATNNKAG 319
QY 65 QTSIAATATSEQASATDAATDOASA-----AEQTQGTASTDTAAQTNNANE 112
DB 320 PLSIGNSIKRKLIINGAGVKNKTDVNVQAEVAVKAKERRITFOGDDNSTDVKIGLDN 379
QY 113 AKWVPTENQGTDEMLAEAKNVAESDISPSDLAK-MSNVQVQDGKYYVQDQGNVK 171
DB 380 TLTIKGAETNALTDDNIGVVKE---ADNSGLVKVLAETLNNLTENV-----TTTLN 428
QY 172 KNFAVSGDKIYFDETGAYKDTFSKYVDADKSSSAVSQNTAFANRRAYSTSAKNEAVD 231
DB 429 ATTVKVGSS---SSSTAELLSLTLTTPQNTQSGSTSKTVYGVNGVYKFTNNAETAAIG 485
QY 232 NYLTADSWYRPKGIKDGKTWESGKDDPRLLMAWMPDPTETKRNVYVNMKNVVG----- 286
DB 486 T-----TRITRDKIGFARDGVD-----EKQAPYLDKKQLKVGSAIT 523

QY 287 IDKTYTAETSOADLTAAAEELVOARIEQKITSENTKWLBEAISAFAVKTOPQWNGESEKPY 346
DB 524 IDNGIDAGNKKITSNLA-----KGSSANDAVTIEQLKAA-----KP-- 558
QY 347 DDHLQNGALLFONQTDLTDPDTQS-----NYRL-LNRTPNTQOTSLSDSFYTNPNPDPLGG 399
DB 559 ---TLNAGAGISVTPFEISVDAKSGNVTAFTYINIGVKTTLSNDSGTS-D-KFSVKS----- 610
QY 400 YDFLLANDVNSPVVQABQLNWLHLYLNFSGIYAND-----ADANPDSIRVDAVDNVDAD 455
DB 611 -----GTNNSLVTAEHL-----ASYLNEVNRATDSALQSFTVKEEDDDAN 651
QY 456 LLIQISSDYKAA-----YCIDK-----NN----- 474
DB 652 AITVAKDTTKNAGAVSILKKGKNGLTVAATKDKGTVTFGLSQDSGLTIGKSLTNGLTV 711
QY 475 KXANNHVS-----VEAWSNDPTPYLHDDGD 500
DB 712 KDTNEQIQVANGIKFTNVGNSPGTGIANTRDKIGFAGSDGAVDTNKPYL--DQD 769
QY 501 NLMMNDNKFRLSMLMSLAKPLDKRGLNPLIHNSLVDRVDDREVE---TVPSYSFARAH 557
DB 770 KLVQGNVKITNTGINAGGKAI--TGLSPTL-PSIADQ--SSRNIELGNTIQDKD--KSN 821
QY 558 DSEVQDIIRDIKAEINPNSFGYSFTQBEIEQA-----FKIYNEDLKTKDKKYTHYVNP 611
DB 822 AASINDILNTGPNLKNNNPIDFVSTYDIVDFANGNATTATVTHDTFANKTSKVYVDNVND 881
QY 612 LSVYTLTLTNKGSIPRVYVYGDMEFTDDGOYMAKNTVNYDAIESLLKARMKVSVGQAMQNYQ 671
DB 882 -DTHIHLTG-----TDDNKKLGKVTIKLN-----KTSANGNTATNFN 917
QY 672 IGNG--EILTSVRYGKGALK-----QSDKGDATRTTSGVVMGMGNPNSLDGKVVVALN 723
DB 918 VNSDEEDALVNAKDAENLNTLAKEIHTTKGTADTALQFTTVKVKVDENNADDAN--AIT 975
QY 724 MGAHANQBYRALMVSTKGVATYATDADAKAGLVKRTDENGILYF----- 770
DB 976 VGGKANNQVNTLTLLKGEN-----GLNIKTKNGTVTFGINTTSGLKAGKS 1021
QY 771 -LNDDLKGVANPQVSGFLQVWVP-----VGAADDQDIRVAASDT--ASTDGKS 815
DB 1022 TLNDGGLSINKNPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTTRITREIGFTGNG-S 1080
QY 816 LHQDAAMDSTRVMEGFSNFPQSPATKEEYTNV-----VIANNVDKVFWSGITDEM-APQY 870
DB 1081 LDKSKPHLSK--DGIN-----AGGKITNIQSGEIAQNSHDAVTGGKIYDLKTELENK 1131
QY 871 VSST-----DGOFLDSVIQNGYAFTRDYDLGMSK-----ANKYGTADQLVK 911
DB 1132 ISSTAKTAQNSLHEFSVADEQGNNTFVSNPYS---SYDTSKTSDDVTTFAGENGITTKVKN 1188
QY 912 AIKAL---HAKGLKVMADWVPQWYTFPKOEVVVTVTRTDKFGKPIAGSQINHSLYVTD-- 966
DB 1189 GVVRVIGIDTKGL-----TTPK-----LTVGNNGKGIVIDSQNGQNTITGLS 1231
QY 967 -----TKSSGDDYQAKYGGAFDELKEYPELFTKKQI-----STGQAIIDPSVKIKQWS 1015
DB 1232 NTLANVTNDKGSVRTTEQGNIIKEDKTRAASIVDVLSAGENLQNGEAVD---FVSTYD 1288
QY 1016 AKYFNGSNILGRGADVVLSDQVSNKYF--NVASDRTLFLPSSLLGKVVESGIRYDGGK--- 1070
DB 1289 TVNFADGNATTAKTYYDDTSKTSKVYVDYVNDVDDTTIEVKDKKLG-VKTTLTSTGTGANK 1347
QY 1071 TYVNSSATGDQ-VKASFI-----TEAGNLYYFGKQGYWVTGAQTINGANYFFLENGTALR 1124
DB 1348 FALSNOATGDALVKASDIIVAHLNLTLSGDI-----QTAGAS-----QANN 1387
QY 1125 NTIYTDAGNSHYIANDGKRYENGYYQFGNDMRYFKDGNMVAVLTTVDGNNVYFDKDGQV 1184
DB 1388 SAGYVDADGNKVIYDST---DNKYYQAKN-----DGTV---DKTKEV 1423

Qy	1149	-YQFGNDWRYPKD-----GNMAVGLTTVD-----GNVQYFDKD-GV-----QAKD	1187
Db	1055	NYVRTNDGLAFNDASAQGVGATAIGNSVAKGSSVAIGQGSYSYDVTGIALGSSSVSS	1114
Qy	1188	KII-----VTRDGKVRYPDONGNAATWTFADTGHWYILG-XDGVAVTGAQIVGK	1238
Db	1115	RVIAGSRDTSITENGWVIGYDITDGLLGLALSIGDDGKVRQIINVADGSEAHDAVTVRQ	1174
Qy	1239	QK-----LYPEANGQOVKGFVTSDEGKLYFYDVDSGD-----MMWDTFIEDKAG	1283
Db	1175	LQNAIGAVATTPTKYFPHANSTEEDSLAVGTDLSLAKAKTIWNGDKGIGYGVAVDANAL	1234
Qy	1284	NWPLYLKGDAV-----TGAQITIRGOKLYF-----KANGQOVKG	1317
Db	1235	NGIAGSNAQVIHNSIAGNGSTTTTGAQNTYATYAMDAPONSVEFSVGSADGQRIIT	1394
Qy	1318	DIVKG---TD---GKIRYDAKSGEQV-----FNKTVKAAADGKTYVIGNDGVAVDPSPV	1365
Db	1295	NVAAGSADTDVAVNVQQLKVTDAAQVSYQNTQSTITLNDNRVTLSRVTNIEN---GIGDIVT	1351
Qy	1366	KGQT--FK-----DASGALRFPYLNKQLVGTGSGWYETANHWDVYIOSGKALTGEOTI--	1415
Db	1352	TGSTKYFKTNTDGVDAASA---QKQDSVAIGSGIAAADNS-VALGTGSVATEENTISV	1405
Qy	1416	---NGOHLFYKEDGHQVK---GQLVTGTDGKVRYYDANSBGDQAFNKSVTVNGKTYIF	1466
Db	1406	GSSTNQRRIITNVAAGKKNATDANVVAQLKSEAGGVR-YDTKAGSDIDYSNITLGG-----	1459
Qy	1467	GNQDTAQTAGNPKG-----QTTFKDGSDIRFYSMEGOL-----	1498
Db	1460	GNCGTTRISNVAGSVNNNDVVNYAQLKQSVOETKQYTDQRWEMDNKLSKTESKLSGGIA	1519
Qy	1499	-----VTGSGWYE-----NAQOGWLTVKNG	1518
Db	1520	SAMAVITGLPQAYTFPGMASIGGGTYNGESAVALGVSMVSGANGRWYVKLG	1570

RESULT 11

```

US-10-793-626-2964
; Sequence 2964, Application US/10793626
; Publication No. US2005025478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid
; US-10-793-626-2964

```

Qy	80	TDAATDQASAAEQTOGTTAS-----TDTAAQTNTTNAEAKWVP-----	117
Db	2466	AEAKSDNGOSIEHLQGLTOSQDKQHQHDLINQAOTKQGVDDIVNNSKQLONSMNMQLOQIVN	2526
Qy	118	---TENENQOCTDB-----MLAEAKGVATAESDSIPSLAK-----MSNVKQVD	158
Db	2526	NDNTVKQNSPFINEDSSQDAYNHAQAAKDLITAH-----PTIMDKKNQIDQAIENIKQAL	2581
Qy	159	GKYYYYQDQGNVKKNFAVSGDKIYYFDE--TGAYKOT-----SKVDADKSSAV	206
Db	2582	NDLHGSNKLSKXK---EASEQLQNLNSLTNGOKOTILNHIFSAPTRSQV--CEKIASAK	2636
Qy	207	SONATIFA-----ANNPAYSTSAPF-----EAVDNYLTADSWTRPRKSIILKDGKTWTESGK	257
Db	2637	QLANNTKALRDSIADNNEILQSSKYFNEDEQQNAYNAQV--KAKNIINDQPT--PVMAN	2693
Qy	258	DDREPLLMWMPDETETKRNYYN--YMKKVGIDIKTYTAETSOADLTAAAEIQAAR-----	313
Db	2694	DETOGVL-----NEVKQTKDNLHGQOKLANDKT-----DQAOTLNALYNLQAQRNLET	2743
Qy	314	KITSENNKTWLRBAISAFVKVPTOPQWNGESEKPYDDHLQ--NGALL-----F	357
Db	2744	KVQNSNS---RPEQVKVQLANQLN--DAMKKLLDALTGNDAIKQTSYINEDTSQOVNF	2798
Qy	358	DNQTDLTPTQOSVYRLLNRTPTNQTSGLSORFTYNNPNDPLGGVDFLLANDVONSVPV---	414
Db	2799	DEYTRGKNIVABQTPNPNMSPTN--INTIADKITEAKND--LHGVOKELEQAQOQSINTINQM	2856
Qy	415	---VQAEQLNW-----LHYLLNFGSIYANDADANPDSIRVDADVNDVADLLQI	459
Db	2857	TGLNQAKQELAQEIQOTRSEVHOI-----NKAQALNDSMMTLRQSITDEHEVKQ	2909
Qy	460	SSDYLKAAAYGIDKXNKNNANNHVSIVBEASDNDDTPYLHDDGLNMNMNDKFRLLSMLWSLAK	519
Db	2910	TSYINETVGNQYVYNAVDRVKQIINQTSNPT-----MN-----	2944
Qy	520	PLD--KRSGLN-----PLIINSLVDREVDRETVETVPSYSFARAHSEVQDIIRDIKABI	573
Db	2945	PLEVERATSNVKTSDALHG--ERELENDK---NSKTFVANHLDNLNQAQKEALTHEI	2997
Qy	574	NPNISFGVSPQEBIEQAFIYNE-----DLKK--TDKKYTHYVPLSYTLTLTNKGSIRP	626
Db	2998	EQATI-----VSQVNNIYNKAKALNNDMKGLKDIVAQQDNVRSNNYI--NEDSTPQ	3047
Qy	627	VYXGDMFTDDG---QYMAKTVNYDAIESL---LKARMKYVSGQAMQVQIENGIBILTS	680
Db	3048	NMYNDTINHAQSIIDQVAPNTMWSHDEIENAINIKHAINALDGEHKLQAQKENVALLINS	3107
Qy	681	VRYKGALKQSDK---GDATRTTSGVGVVGNQPNFSLDGKVVALMNGAHAHQEYRALM	737
Db	3108	LNDLNAQPDARNLVNEAQTREKVAEQLQSAQ-----ALNDAMKHLRNSIQNGS	3157
Qy	738	VSTKDGVAIYATDADAKAGLVKRTDENGLYFLND-----DLKGVANPOV	783
Db	3158	SVFQESKYINASDAKKEQYNHAREVEN---IINEQHPTLDKEIKOLTDAVNOQANNDL	3213
Qy	784	SGFLQVWVPVGAADDQDIR-----VAASDFTASTDGSLSLHOD	819
Db	3214	NG-----VELLDADKQNAHQSIPTLMHLNQAQONALNEKINNAVTRAKVAAIIGQAKIJD	3268
Qy	820	AAMDS-----RVMPEGFSNF--QSPATKBEYTNVV-----IANNVDKFVSWGITDPE	865
Db	3269	HAMENLEESIKQKEQVKQSSVINEBPDVQOETYNNAVHVHTILNQTVPNPTLS--IEDIE	3326
Qy	866	MAPQYVSSDTGGFLDSVIONGVAFTRDRLDGLMSKANKYGTADQLVKAIKALHAKGLKWA	925
Db	3327	HATNEYVNAQKQL-----RGKQKLYQTID-----LADKELSKLD	3360
Qy	926	DWVPDQMYTFPKQEVVYVTRTDKFGKPIAGSOINHSLYVTDTKSSGDDYQAKYGGAFIDE	985
Db	3361	DLTSQOSSISINQIYTAKTREVAQAIEKAKSLNHAM-----KALKYI--KNADKVLDS	3413
Qy	986	---LKEYPELFTKKQI-----STGQAI DPSP-----K	1010

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-022-562-228

Query Match      3.1%; Score 257; DB 7; Length 396;
Best Local Similarity 25.2%; Pred. No. 3.7e-07;
Matches 131; Conservative 64; Mismatches 151; Indels 174; Gaps 30;

QY 969 SSGDDYQAKYGGAFADELKEKYPELTTKKOISTGOAIDPSVKIK--QWSAKYFNGSNI-- 1024
Db 25 SSSFYKWSSTEGS--DFILVYLESNKK-----ILQIRIKIGLSLTKSFNKSIDF 75
QY 1025 -----LGRGADYVLSDOVSNNKYFNVASDTLFLPSSLLG-KVYESGIRYDGKGIYNNSAT 1078
Db 76 KDIKKLSLG--YIMS--NFKSFENSE--LDRDHLGFKII-----DNKTYVYDEAS- 120
QY 1079 GDOVKASFITEAGNLYYFGK-DGYMVTGAQTINGANYFPLENGTALRNTIYTDAGNSHY 1137
Db 121 --KLVKGLININNSLFYFDPIESNLVTGMQTINGKKYIF-DINTGAASTSYKIING-KHF 176
QY 1138 YANDGKRYENGYYQQFG-----NDWRYFKDGNMAVGLTTVDGNVQYFDKDGVOAKDKIIVT 1192
Db 177 YFNN-----NGVMQLGVFKGPDGFYFAPAN-----TONNNI---EGQAIYVQSKFL-T 221
QY 1193 RDGKRYRYFDQHNNAATNTFIADKTGHWYLYLGKGVAVTGAQTVGKQKLYFANGQOVKG 1252
Db 222 LNKKKYIFD-----NDSKAVTGMQITDGKKYFNLNTAEAAAT 258
QY 1253 DFTVSDGKLYPYDVSDGDMWTDFTIEDKAGNWFYLGKGAATGAAVTGQATIRGQKLYFKANG 1312
Db 259 GQQTID-GRKKYFN-----TNTSI-----ASTGYTIINGKHFYFNTDG 295
QY 1313 QQVKGDIVKGTQDKIRYDYDAKSQEVFNKTVKAADGKTVYIGNDGVAVDPVSVVKGQTFKD 1372
Db 296 IMQIG-VFKGPG-PEYF-----APANT-----DANNIEGO----- 324
QY 1373 ASGALRFYNLKQLVTGSGWYETANHDWYI--QSGKALTGEQTINGOHLKFKEDGHQVK 1430
Db 325 ---AIRYQNRFLYL-----HDNIYFNGNSKAVTGMQTINGNVVYFMPD----- 365
QY 1431 GQLVTGTGDKVRYDANSQDQAFNKSVTVNGKTYIFGNDG 1470
Db 366 -----TAMAAAGGLFEIDGVIYFEGVDG 388

RESULT 13
US-10-873-528-17
; Sequence 17, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-17

3414 SRFINEDQPEKAYQQAIIINHVSIIHQTNPEMDPTVINSIITHELETAQNHLHGQKLAH 3473
QY 1011 IKQWSAKYFNGSNILGRGADYVLSDOVSNKYFNVASDTLFLPSSLLGKVVESGIRYDGKG 1070
Db 3474 AKQDAANVINGL-----IHLNVAQREVMIWT----- 3499
QY 1071 YIYNSATGDQVKASFITEAGNLYYFGKGYMVTGAQTINGANYFPLENGTALRNTIYTD 1130
Db 3500 ---NTNATTREKVAKNLNDNAQAL-----DKAMET-----LQGVAAHKNILND 3539
QY 1131 AQGNSHYYANDGKRYENGYYQQFGNDWRYFKDGNMAVGLTTVDGNVQY---PDKDGVQAKD 1187
Db 3540 SK-----YLNEDSKYQQQYD-----RVTADEAQLNQTNPTELPYKVDIVKDNVLANE 3588
QY 1188 KIIIVTRDGRVYFDQHNNAATNTFIADKTGHWYLYG-----KQGVAVTGAQTVGKQK 1240
Db 3589 KILFGAE-KLSY-DKSNAN-----DEIKHMYLNNNAQKOSIKDMISHAALRTEVKQL 3638
QY 1241 LYFEANGQQVKGDFVTSDEGKLYFYDVSDGDMWTDFTIEDKAGNWFYLGKGAATVGAQT 1300
Db 3639 L-----QQAK-----TLDEAMKSLD-KTQVWITDTLP-----NYTEASEDKKEKVDQTV 3683
QY 1301 IRGQKLYFKANGQOVKGDIVK-----GTGDKIRYDYAK----- 1333
Db 3684 SHAQAIIDKINGSNVLSDQVRQALEQLTQASENLGDGQVRVEAKVHANOTIDQLTHLNSL 3743
QY 1334 ----SSEQVFNKT-----KKAADGKTVYIGNDGV 1358
Db 3744 QQTAKESVKNATKLEIATASNALALNKMVKLEQFINHADSINSDNYRQADDKII 3803
QY 1359 AVDPSPVVKQ-----TFKDGASGALRFYNLKQOLV---TGSQWYETANH---DWYITQ 1404
Db 3804 AYDDALEHGQDIQKSNATQNEAKQALQ-----QLINAETSLNGFERLNHARPRALEYIK 3857
QY 1405 S-----GKALTGEQTINGOHLKFKEDG---HQVKQOLVTG-----TDGKV 1441
Db 3858 SLEKINNAQSALEDKVTQSHDLLEHLV--NEGNTLNDIMGELANAVNNYAPTAKSI 3915
QY 1442 RYDA-----NSGDOAFNKSVTVNGKTYFVGNDGTAQ----- 1473
Db 3916 NYINADNLKDNFTQAINARDALNT---OQQNLDFNAIDTFKDDIFKTKDALNGIERL 3972
QY 1474 TAGNPKGQTFKQSGDIRFY-----SMEGQ 1497
Db 3973 TAAKSAEKELID--SLKFINKAQFTHANDEIMNTNSIAQLSRIVNOAFDLNDAMKSLRDE 4030
QY 1498 L-----VTGSGHYENAO-----GQWLYVXNGKVLTLGLQTVGSRQVYFDE- 1536
Db 4031 LNNQAHFPVQASSNYINSDLEDLKQOQFDHALSNARKVLAKENGKRLDEIQIEGLKQVIEDTK 4090
QY 1537 ---NGIQ---AKGKAVR 1547
Db 4091 DALNGIQRLSKAKAKAIQ 4108

RESULT 12
US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
```


Db 123 TNESTPKTTEA-----PTVNE-----ESIAETPKTSTTQODSTEKNPSLK 164
Qy 153 NYKQVDGKYYYDQDGNVKNFAVSGDKIYYFDETGAYKDTSKVDADKSSSAVSQNAI 212
Db 165 D-----NLNSSSTTSKESKTDEHSTKQAQMSINKSNLNDTNDSPQSEKTS 209
Qy 213 FRANNRAYSTSAKNFAVDNYLTADSWYRPK-----SILKDGKTWTESGKDD 259
Db 210 SQANNDSTNQAPSQKLSKPFSEQVKYTKFNDEPTQDVEHTTKTLKTPSISTDSSVND 269
Qy 260 FRPLLMAWPDTEKENVYNNKVVGIDKTTVAETSQA-----DLTAAAEVLQARIQOK 314
Db 270 -----KODYTRSAVASLGVDNSNETEALTNVRONLDLKAAS-----REQINEA 312
Qy 315 ITSENNTKWLREAIASFVKTPQPMNGESEKPPDDHLQNGALLFDNQDTLTPDTQSNYR-- 372
Db 313 IIAEALKK-----DFSNP-DYGVDTPLALNTSQSKNSPHKSPRWN 353
Qy 373 LLNRTPTNQTGLSDRFTYNPNDPLGGYDFLLANDVDNSNPVVQABQLAWLHVLNPGSI 432
Db 354 LMSLAAEPNSGK-----NVND-----KVKITNPTLS-----LNKSN 385
Qy 433 YAND-----ADANFOSIRVDAVDNVDADLLQISSDYLKAAGYDKKNKNANNHVSIVEAW 487
Db 386 HANNVIWPTSNOFN-----LKANYELDDSIK----- 412
Qy 488 SOND-----PYLHDDGDNLMNMKNFRLSMLSLAKPLDKRSGLNPLIHNLSLVREYD 541
Db 413 -EGDTFTIKYGVYIRPGGLEPAIKTQLR-----SKDG-----SIVANGVY 452
Qy 542 DREVEYPSYSFARAHSEVDIIRIIKABINPNSFGYSFTQBEIEQAQFKYNEDLKKT 601
Db 453 DKTNTTT-TYTFNTYVD-QYQNI-----TGSF-----DLIAT 482
Qy 602 DKYKT-----HYNVPLSYTLLTNKSGI PRVYVGMFTDDQYMAKNTVNYDA-IESLLKA 656
Db 483 PKRETAIKONQNYPEVITI-----ANVVKKDFIVDYGKKNKONTTAAVANVDNVNKK 535
Qy 657 RMKYVSGQAMQYQIGNGEILTSVRYGK--GALKQSDKGDAIT-----RTSGVG 704
Db 536 HNEVVYLNQNNQPKY--AKYFSTVNGKGFIPGEVKVYEVTDINAMVDSFNPDLNSSNVK 593
Qy 705 VVNGN-QPNFSLDGKVVALNM-----GAAHAN 730
Db 594 DVTSQFTPKVSDGTRVDINFARSMANGKKYIVTQAVRPTGNGVNYTEYWLTRDGTNTN 653
Qy 731 QEYRALMVSTKDGVAITYATDADASKAGLVKRTDENGLYFLNDDLKGVANPQVSGFLQW 790
Db 654 DFR-----GKSTTVTLNGSSTAQGD--NPTYSLGDYVWLDKNKNGVQDDDEKGLAGVY 707
Qy 791 VPVGAADDQDIRVAASDTASTDGKSLHQDAAMDSTRVMEFGSNFQS-----FATKEEY 844
Db 708 VTLKDSNNRELQ-----RVTTDQSGHYQ-----FDNLQNGTVTVEFAL-PDNY 749
Qy 845 TNVVIANNVDKFSW-----GITDFEAPQYVSTGQFLDSVIQNGYAFTDRYDLG--- 896
Db 750 TSPANNSTNDATSDSGERDGRKVVVAKGTINNADNMTVDT-----GFYLTPKYNGDYV 805
Qy 897 MSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQYTPPKQEVVTVTRTDKFGKPIAGS 956
Db 806 WEDTNKDGIGDNEKISN-----VKVTLKNKNGDTIG-- 838
Qy 957 QINHSLYVTDTSKSGD-DYQAKYGGAFDLDELKEKYPELFTKKQISTGOAIDPSVKIKQWS 1015
Db 839 -----TTTDDSGNGKVEFTGLENGDYTIEFET--PEGYPTTKQNSGD-----E 879
Qy 1016 AKYFNG--SNILGRGADYVLSDQVSNKYFNVASDTLFLPSSLLGKVVESGIRYDGKY-- 1071
Db 880 GKDSNGTKTTVTVKADAD-----NKTIDSGF-----YKP 907
Qy 1072 IYNSSATGQOVKASFITEAGNLVYFGKDGVMVTGAOTINGANYFFLE-NGTALRNTIYTD 1130
Db 908 IYN---LGD-----YVWEDTN-----KDGIQDDSEKISGVKVTLDKKNGNAI-GTTTTD 953

Qy 1131 AQGNSHYYANDKRYENGVOQFNGDNWRYFKDGNMAVGLTTVDGNVOYFDKDGVOAKDKIL 1190
Db 954 ASG--HY---QPKGLENGSY-----TVE----- 971
Qy 1191 VTRDGKVRVFDQHNHNAATNTFIADKTGHWYVLGKDGVAVTGAQTVGKQKLYFEANGQOV 1250
Db 972 -----FETPSGYTPTKA-----NSGQDITVDSNGITTTGI-----ING--- 1004
Qy 1251 KGDFTVTSDEGKLYFYDVIDS---GD-MWTDFTIEDKAGNWFYLGKDGAAVTGAQTIRGOKL 1306
Db 1005 -ADNLITDSG---FYKTPKYSVGDYVWEDT-----NKDGIQDDNEKISGVKV 1048
Qy 1307 YPKANGQQVKGIIVKGT-----DGKIRYDYDAKSGBQVFNKTVKAADGKTVVIGNDGVAVDP 1362
Db 1049 TLK-----DEKGNIIISTTTTDDENGKYQFONLDSNGYIIH--PEKPEGMTOTTANS 1097
Qy 1363 SVVKGQTPKDSG 1375
Db 1098 -----NDDEKADG 1106

Search completed: February 11, 2006, 20:58:23
Job time : 22.048 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 32.8183 seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-37

Perfect score: 8349

Sequence: 1 MEKNVRFKHKVKRWLTLS.....VGYQYYFGNDGARIYRGWN 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_80.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8256	98.9	1592	2 A38175	glucosyltransferas
2	4580	54.9	1475	2 B33135	gt5B protein precu
3	4214.5	50.5	1375	2 J70345	dextranucrase [EC
4	3599	43.1	1431	2 A45866	dextranucrase [EC
5	3325.5	39.8	1518	2 A4811	glucosyltransferas
6	3245	38.9	1577	2 T30858	glucosyltransferas
7	3214.5	38.5	1599	2 S22737	glucosyltransferas
8	3128.5	37.5	1449	2 T30857	glucosyltransferas
9	3114.5	37.3	1449	2 T30552	glucosyltransferas
10	3088	37.0	1508	2 T31098	probable dextranu
11	3031	36.3	1365	2 A41483	glucosyltransferas
12	3021	36.2	1290	2 JCS473	dextranucrase [EC
13	678	8.1	2817	2 B97033	uncharacterized pr
14	518	6.2	563	2 A37184	glucan-binding pro
15	513	6.1	2710	2 A37052	toxin A - Clostrid
16	435.5	5.2	648	2 S10869	enterotoxin A - Cl
17	416	5.0	2178	2 S55805	alpha-toxin - Clo
18	400.5	4.8	2364	2 I40884	cytotoxin L - Clo
19	388	4.6	2366	2 S10317	toxin B - Clostrid
20	381	4.6	2367	2 S70172	toxin B - Clostrid
21	329	3.9	1806	2 AF1717	probable peptidogl
22	328.5	3.9	2334	2 S32920	cell wall-associat
23	311	3.7	1385	2 D98824	hypothetical prote
24	299.5	3.6	1829	2 E81086	iron-regulated pro
25	298.5	3.6	1946	2 AE1449	hypothetical prote
26	291.5	3.5	1829	2 S35027	cytotoxin RTX homo
27	291	3.5	1315	2 T28679	fibrinogen-binding
28	284.5	3.4	1463	2 T30290	AAS surface protei
29	277	3.3	2167	2 AF1489	cell wall-associat

30	272.5	3.3	1612	2 AB1347	probable peptidogl
31	272	3.3	1946	2 JC6032	lactocepin (EC 3.4
32	268.5	3.2	721	2 C97980	endo-beta-N-acetyl
33	265.5	3.2	2059	2 D82671	surface protein XF
34	265	3.2	1588	2 A86036	probable adhesin Z
35	265	3.2	1588	2 H91188	probable adhesin E
36	264.5	3.2	1999	2 AB2018	hypothetical prote
37	257	3.1	13055	2 T16580	hypothetical prote
38	254.5	3.0	658	2 E95111	endo-beta-N-acetyl
39	254.5	3.0	6713	2 B89921	hypothetical prote
40	253.5	3.0	329	2 A55221	dextranase inhibit
41	253.5	3.0	1302	2 C81182	iron-regulated pro
42	251.5	3.0	1383	2 G86643	hypothetical prote
43	249.5	3.0	1959	2 AG1085	hypothetical prote
44	248	3.0	619	2 A97887	surface protein ps
45	248	3.0	619	2 A41971	surface protein ps

ALIGNMENTS

RESULT 1

A38175

Glucosyltransferase precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C>Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004

C:Accession: A38175

R:Abo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

J. Bacteriol. 173, 989-996, 1991

A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus

A:Reference number: A38175; MUID:91123227; PMID:1704006

A:Accession: A38175

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1592 <ASO>

A:Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID

F:1093-1112/Domain: cpl repeat homology <CP1>

F:1222-1241/Domain: cpl repeat homology <CP2>

F:1287-1306/Domain: cpl repeat homology <CP3>

F:1330-1351/Domain: cpl repeat homology <CP4>

F:1352-1371/Domain: cpl repeat homology <CP5>

F:1402-1420/Domain: cpl repeat homology <CP6>

F:1465-1484/Domain: cpl repeat homology <CP7>

F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match

Best Local Similarity 98.9%; Score 8256; DB 2; Length 1592;

Matches 1574; Conservative 6; Mismatches 10; Indels 2; Gaps 1;

QY	1	MEKNVRFKHKVKRWLTLSVASATMLASALCASVASADTDITASDSDSNOAVVTGDDTTNN	60
Db	1	MEKNVRFKHKVKRWLTLSVASATMLASALCASVASADTDITASDSDSNOAVVTGDDTTNN	60
QY	61	QATDQTSIAATATSSQASASTDAATDQASAAEQGTGTTASTDTAAQTNNANEAKVPTEN	120
Db	61	QATDQTSIAATATSSQASASTDAATDQASAAEQGTGTTASTDTAAQTNNANEAKVPTEN	120
QY	121	ENQGTDEMLAEAKNVATAESDSIPSLAKMSNVQVKGKYYQDQGNVKNFAVSVGD	180
Db	121	ENQGTDEMLAEAKNVATAESDSIPSLAKMSNVQVKGKYYQDQGNVKNFAVSVGD	180
QY	181	KIYYFDETGAYKDTSKVDADKSSSAVSONATIFAANNRAYSTSAKNFEADVNTLADSWY	240
Db	181	KIYYFDETGAYKDTSKVDADKSSSAVSONATIFAANNRAYSTSAKNFEADVNTLADSWY	240
QY	241	RPKSLKDGKWTESGKDDFRPLLMAMPDPTETKENYNNMKVGVGIDKTYTAETSQADL	300
Db	241	RPKSLKDGKWTESGKDDFRPLLMAMPDPTETKENYNNMKVGVGIDKTYTAETSQADL	300
QY	301	TAAELVQARIEQKITSENNTKWLREAIASFVKTPQWNGESEKPYDDHLQNGALLFDNQ	360
Db	301	TAAELVQARIEQKITSENNTKWLREAIASFVKTPQWNGESEKPYDDHLQNGALLFDNQ	360

QY 361 TDLTPDTOSNYELLNRTPTNQTGSLDSBFTYVNPNDPLGQYDFLLANDVDNSNPVVOABOL 420
DB 361 TDLTPDTOSNYELLNRTPTNQTGSLDSBFTYVNPNDPLGQYDFLLANDVDNSNPVVOABOL 420
QY 421 NWLHYLLNFGSIYANDADANFDSIRVDADVNDVADLLOI SSDYLKAAAGIDKNNKANNH 480
DB 421 NWLHYLLNFGSIYANDADANFDSIRVDADENDVADLOI SSDYLKAAAGIDKNNKANNH 480
QY 481 VSIWEASNDTPYLLHDDGDNLMNNDKFRLSMLWSLAKPLDKRSLGNPLIHNSLVDRV 540
DB 481 VSIWEASNDTPYLLHDDGDNLMNNDKFRLSMLWSLAKPTDVRSLGNPLIHNSLVDRV 540
QY 541 DDRETVTPSYSFARAHSEVODIIRDIKABINPNSFGYSTOSEIEQAQFIYNEDLKK 600
DB 541 DDRETVTPSYSFARAHSEVODIIRDIKABINPNSFGYSTOBEIDQAQFIYNEDLKK 600
QY 601 TDKKTYTHNVPLSYLLLTNKGSIIPRVYVGMFTDDGQYMAKNTVNYDAIESLLKARMKY 660
DB 601 SDKKYTHNVPLSYLLLTNKGSIIPRVYVGMFTDDGQYMAKNTVNYDAIESLLKARMKY 660
QY 661 VSGQAMQNYQIENGIEILTSVRYGKALKQSDKGDATRTSGVGVMMGNQPNFSLDGKV 720
DB 661 VAGQAMQNYQIENGIEILTSVRYGKALKQSDKGDATRTSGVGVMMGNQPNFSLDGKV 720
QY 721 ALNWGAHANQBYRALMYSYTDGVAATYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
DB 721 ALNWGAHANQBYRALMYSYTDGVAATYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
QY 781 PQVSGFLOVPVGAADQDDIRVAASDTASTDGKSLHODAAWDMVMEFEGSNFOSFATK 840
DB 781 PQVSGFLOVPVGAADQDDIRVAASDTASTDGKSLHODAAWDMVMEFEGSNFOSFATK 840
QY 841 EEEYTNVVIANNVDFXVSGWITDFEMAPQYVSSSTDGQFLDSYIONGYAFTRIDYDLGMSKA 900
DB 841 EEEYTNVVIANNVDFXVSGWITDFEMAPQYVSSSTDGQFQDSYIONGYAFTRIDYDLGMSKA 900
QY 901 NKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFPKQEVVTVTRTDKFGKPIAGSQINH 960
DB 901 NKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFPKQEVVTVTRTDKFGKPIAGSQINH 960
QY 961 SLYVTDTKSSGDDYQAKYGGAFDLDELKEKYPBLFTKKOISTGQALDPSVKIKQWAKYFN 1020
DB 961 SLYVTDTKSSGDDYQAKYGGAFDLDELKEKYPBLFTKKQWSTGQALDPSVKIKQWAKYFN 1020
QY 1021 GSNILRGADYVLSDOVSNKYPNVASDTLFLPSSLLGKVVESGIRYDGKGIYNSSATGD 1080
DB 1021 GSNILRGADYVLSDOVSNKYPNVASDTLFLPSSLLGKVVESGIRYDGKGIYNSSATGD 1080
QY 1081 QVKASFITEAGNLXYFGKDGVMYVTAQOTINGANYEFLENGTALRNTIYTDAGNSHYAN 1140
DB 1081 QVKASFITEAGNLXYFGKDGVMYVTAQOTINGANYEFLENGTALRNTIYTDAGNSHYAN 1140
QY 1141 DGKRY - - ENGVOQFGNDWRYFKDGNMAGVLTVDGNVOYFDKDGVOAKDKIIVTRDGKVR 1198
DB 1141 DGKRYENENGVOQFGNDWRYFKDGNMAGVLTVDGNVOYFDKDGVOAKDKIIVTRDGKVR 1200
QY 1199 YFQDHNGAAVNTFTIADKTGHYLYLGKDGVAVTGAQTVGKQKLYPEANGQQVKGDFVTS 1258
DB 1201 YFQDHNGAAVNTFTIADKTGHYLYLGKDGVAVTGAQTVGKQKLYPEANGQQVKGDFVTS 1260
QY 1259 EGKLYFYDVSQDMWTDFTIEDKAGNWFYLGKDGAAVTAQOTIRQOKLYFKANGQQVKG 1318
DB 1261 EGKLYFYDVSQDMWTDFTIEDKAGNWFYLGKDGAAVTAQOTIRQOKLYFKANGQQVKG 1320
QY 1319 IVKGTGDKIRYYDAKSGQVFNKTKAADGKTYIYVGNMGVADVPSPVVKQGTFKDASGALR 1378
DB 1321 IVKGTGDKIRYYDAKSGQVFNKTKAADGKTYIYVGNMGVADVPSPVVKQGTFKDASGALR 1380
QY 1379 FYNLKGQVLVSGWYETANHDWVYIQQSKALTGEOITINGQHLYPKEDGHQVKGQVLVGTGD 1438
DB 1381 FYNLKGQVLVSGWYETANHDWVYIQQSKALTGEOITINGQHLYPKEDGHQVKGQVLVGTGD 1440
QY 1439 GKRYVDANSQDOAFNKSVTVNGKTYIFGNDGTAQTAGNPKGQTFKDGSDIRFIYSMEGQL 1498

DB 1441 GKRYVDANSQDOAFNKSVTVNGKTYIFGNDGTAQTAGNPKGQTFKDGSDIRFIYSMEGQL 1500
QY 1499 VTGSGWYENAGQQLYVKNKGVLTLGLQTVGSQRYVFDENGIOAKGKAVRTSDGKIRYFDE 1558
DB 1501 VTGSGWYENAGQQLYVKNKGVLTLGLQTVGSQRYVFDENGIOAKGKAVRTSDGKIRYFDE 1560
QY 1559 NSGSMITNWKVYVGYQYVYFGNDGARIYRGWN 1590
DB 1561 NSGSMITNWKVYVGYQYVYFGNDGARIYRGWN 1592
RESULT 2
B33135
gtfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C>Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C:Accession: B33135; A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: B33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SHI>
A:Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN:
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171,173-641,'N',643-1475 <SH2>
A:Cross-references: UNIPARC:UPI000017AC5E
A:Experimental source: strain GS-5
F:1036-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>
Query Match 54.9%; Score 4580; DB 2; Length 1475;
Best Local Similarity 56.4%; Fred. No. 1.2e-207;
Matches 896; Conservative 212; Mismatches 338; Indels 144; Gaps 18;
QY 1 MEKVRPFQWVKRKRWVTLVSASATMLASALGASVASADTDTA-----SDDSNQAVVTGQ 56
DB 1 MDKVRVYKLRVKRKRWTVSVASAVMTLTLLSGGLVKADSNESKSIQISNDSNTSVTANE 60
QY 57 TTN--NQATDQTSIAATATSEQSASTDAATQASAAEQTGTTASTDTAAQTTTNANEAK 114
DB 61 ESNVITEATSKQEAASSQTNHTVTSSSTSVWPK-----VVSNPYTVGETASNGEKL- 115
QY 115 WVPTEENOGTDEMLAEA-----KNVATASDIPSD-----LAKMSNVKQVDGKY 161
DB 116 -----NQQTTVTKTSEAAANNISKQTEADTDDVDSNANLQILKLPNVKSIDGKY 169
QY 162 YYYDQGNVKNFVSVGDKIYYPDETGAAYKDTSKVDADKSSSASVSQATIPAAANNRAYS 221
DB 170 YYDNGKVRNFTLIADGKILHDEYGAYTDTSDITVTK--DIVTTSNLYKKYQVYD 227
QY 222 TSAXNFEAVDNYLTADSWYRPKSILKDGKTWTEGKODFRLLMAWVDPDTEPKRYVNYM 281
DB 228 RSAQSFEHVDHYLTAEBSWYRPKYILKDGKTWQSTQEKDPRLLMTWPDQETQRYVNYM 287
QY 282 NKVYIDKTYTAETSOADLTAAELVQARIISQKITSENNTKWLREASAFVKYTOQWNGE 341
DB 288 NAQLGINKYDDTNQNLQNLNIAAATQAKIEAKITLKNKNTDQRTISAFVKYTOQWNGE 347
QY 342 SEKPYDDHLQNGALLFDNQDTLTPDTQSNYRLLNRTPTNQTGSLDSRFTYVNPNDPLGGYD 401
DB 348 SEKPYDDHLQNGAVLYDNEGLKTPYANSYRILNRTPTNQTGKDPRTY--ADNTIGGYE 405

Qy 402 FLLANDVNSNPVQAEQLNLHLHYLLNFGSIYANDADANFDSIRVDVANDVDADLLQISS 461
Db 406 FLLANDVNSNPVQAEQLNLHLFLMNFENIYANDANFDSIRVDVANDVDADLLQIAG 465
Qy 462 DYLKAAAYGIDKNNKNNHVSIVAEWSDNPYLLHDDGDNLMNMNKKFRLSMLSLAKPL 521
Db 466 DYLKAAAGIHKNDKKAANDHLSILEAWSDNPYLLHDDGDNLMNMNKKLRLSLFLAKPL 525
Qy 522 DKRSGLNPLIHNLSLVDREDEVTVPSSYFARAHDSVQDIIRDIKAEINPNPSFGVS 581
Db 526 NQSGMNPPLIINSLVNRDDNAETAAPVSFIRAHDSVQDLIADIKAEINPNVGVYS 585
Qy 582 FTQBEIEQAFKIYNBLLKTKKTHYNVPLSYLLTNKGSIPRVYVYGDFTDDGQYMA 641
Db 586 FTMBEIKKAFIYNKDLATEKCYTHYNTALSYALLTNKGSVPRVYVYGDFTDDGQYMA 645
Qy 642 NKTWNYDAIESLLKARKMYVSGGQAMQNYOINGEILTSVRYGKALKQSKGDATRTTS 701
Db 646 HKTINYEATELLKARIKIVYSGGQAMRNOQVNSEIITSVRYGKALKATDTGDRTRTS 705
Qy 702 GVGVMGNQPNFSLDGK-VVALNMGAAHANOEYRALMVSTKQGVATYATDADASKAGLVK 760
Db 706 GVAVIEGNNPSRLKASDRVVVMGAAHKNQAYRPLLTDTNGIKAYHSDQEA--AGLVR 763
Qy 761 RTDENGVLFLNDDLKGVANPQVSGFLQVWPVPGAADDQDIRVAASDSTA2DGK--SLHQ 818
Db 764 YTNDRGELIFTAADIKGVANPQVSGVLGVWPVVGAA---LIKMPALRLARPHQQWASVHQ 820
Qy 819 DAAMDSRVMPFGFSNFSQFATKEBYTNVVIANNVDKFSVSGITDFEMAPQVVS2DQOF 878
Db 821 NAALDSRVMPFGFSNFSQFATKEBYTNVVIANNVDKFAEWGVTDFENAPQVVS2DGSF 880
Qy 879 LDSVIQNGYAFTRDYLQMSKANKYGTADQLVKAIKALHAKGLKVMADWPDQWYTPKQ 938
Db 881 LDSVIQNGYAFTRDYLQMSKANKYGTADQLVKAIKALHAKGLKVMADWPDQWYTPKQ 940
Qy 939 EVVTVTRTDKFKPIAGSQINHSIYVTDTKSSGDDYQAKYGAFLDELKERYPELFTKKQ 998
Db 941 EVVTVTRVDKYGTPVAGSQIKNTLYVVDGKSSGDDYQAKYGAFLDELKERYPELFTKKQ 1000
Qy 999 ISTQQAIDPSVKIKQWAKYNGSNIILGRGADYVLSQVSNKYFNVA--SDTLFLPSLL 1056
Db 1001 ISTQVMPDPSVKIKQWAKYNGSNIILGRGADYVLSQVSNKYFNVA--SDTLFLPSLL 1060
Qy 1057 GKVVESGIRVDGKVIYNSSATGDOVKASFITEAGNLYYFGDKGVMVTAQTINGVYFF 1116
Db 1061 NQDSQVGFSDGKGVYVYST-SGYQAKNFTFISEGDKWYFPDNGVMVTAQTINGVYFF 1119
Qy 1117 LENGTALRNTIYTTDAQNSHYIYANDGKRYENGYQOF-GNDWRYFKDGNMAYGLTVDGNV 1175
Db 1120 LSNGLQALRDALIKNEDGTIAYYNGDGRYENGYQFMGVRHFPNGEMSVGLTVIDQV 1179
Qy 1176 QYFQKGVQAKDKIIVTRDGKRVYFDQHNAAVNTFTIADKTGHYVYLGKDGVAVTGAQT 1235
Db 1180 QYFDEMGYQAKGKFTVTDGKIRYFQKSGNMVNRNFTIENBEGKWLVLGEDGAATVGSQT 1239
Qy 1236 VGRQKLYFEANGQVKGDFVTSDEGLKYFYVDVDSGDMKMTDFIEDKAGNWFYLGKDAV 1295
Db 1240 INGQHLVFRANGVQVKGFEVTDHGRISYIDGNSGDDQIRNRFVNRNAQGVYFDNNGYAV 1299
Qy 1296 TGAOTIRGQKLYFKANGQVKGDIYVKGTDGKIRYVYDAKSGEQVFNKTKAADGKTYVIGN 1355
Db 1300 TGARTINGQLLYFRANGVQVKGFEVTDYGRISYIDGNSGDDQIRNRFVNRNAQGVYFDN 1359
Qy 1356 DGVAVDPVSVKGVQTFKDSAGLARFYNILKGQVLVTGSGWYETANHDWVYIQSGKALTGEQTI 1415
Db 1360 NGYAV-----TGARTI 1370
Qy 1416 NGQHLVFKEDGHQVKGQVLVTGTDGKRVYVYDANSGDQAFNKSVTYNGKTYYPGNDGTAQTA 1475
Db 1371 NGQHLVFRANGVQVKGFEVTDHGRISYIDGNSGDDQIRNRFV----- 1412

Qy 1476 GNPKGQTFKDGSDIRFYSGMEGLVGTSGWYENAOQWLYV-KNGKVLTLGTLQTVGSRVYF 1534
Db 1413 -----RNAQCGWFFYDNGVAVTGARTINGQHLVYF 1442
Qy 1535 DENGLOAKGKAVRTSDGKIRYFDENSGSMI 1564
Db 1443 RANGVQVKGFEVTDYGRISYIDANSGBRV 1472
RESULT 3
JT0345
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C:Accession: JT0345; C31135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: JT0345; MUID:89137980; PMID:2976010
A:Accession: JT0345
A:Molecule type: DNA
A:Residues: 1-1375 <UED>
A:Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515
A:Experimental source: GS-5
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: C33135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: UNIPARC:UPI000014825D; GB:M17361
C:Genetics:
A:Gene: gtfC
C:Function:
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
C:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-1375/Product: Glucosyltransferase #status predicted <MAT>
F:1126-1145/Domain: cpl repeat homology <CP1>
F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>
Query Match 50.5%; Score 4214.5; DB 2; Length 1375;
Best Local Similarity 60.3%; Pred. No. 1.7e-190;
Matches 811; Conservative 186; Mismatches 296; Indels 53; Gaps 14;
Qy 1 MEKNVRFPMHKVKRWVTLVSASATMLASGASVASADTDTASDSSNOAVVTGDT--T 58
Db 1 MEKNVRFKLKVKRWVTVSASAVTLTSLGSLVKAAD---STDROQAVTESQASLVT 57
Qy 59 NNOATDQTSIA---ATATSEQASDAATDOASAEQQTGTAST-DTAAQTTTANBA- 113
Db 58 TSEAAKETLTATDSTATSQPTATVTDNVSTTNQSTNTTANTANFVVPVPTTSEQAK 117
Qy 114 -----KWVPTENEN---QGFTDEM-----LAAKAVATAESDS 143
Db 118 TDNSDKIITTSKAVNRLTATGKFVPANNNTAHPKTVTDKIVPIKPIKGLKQPSLSQDD 177
Qy 144 IPSDLAKMNVKQVGDGKYVYDQGNVKQNFVSGDKIYYFDETGAYKDTSKVDADKSS 203
Db 178 IAA-LGNVKMIRKNGKYYKEDGTQKNTALNKGKTFPFFDETGALSNNT-LPSKGN 235
Qy 204 SAVSQNATI FAANNRAYSTSAKNFEAVDNYLTADSWYRPKSIKDKGTWTSEGDFFRL 263
Db 236 ITNNDNTNSFAQYNQVSTDVANFEHVDHYLTASWYRPKYILKDKGTWTSTQTEKDFRL 295
Qy 264 LMAWVPTETKRNVTNPNKVVIGIDKTYTATSTQADLTAAAEVQARIEQKITSNNTKW 323
Db 296 LMTWVDPQETQRQVYVNMNAQLGIHQTYNTATSPQLNLAAQTIQTKEEBKITAENKTNW 355

```
Qy 324 LREAIASFVKTPQWNGSEKPYDDHLQNGALLFDNQTDLTPDTQSNVRLNRTPTNQTG 383
Db 356 LQRTISAFVKTSQAWNSSEKPFDDHLQKGALLYSNNKLSQANSYRIILNRTPTNQTG 415
Qy 384 SLDSRFTYNPNDPLGGYDFLLANDVDNSNPVQAEQLNWLHLYLNFSGSIYANDADANFDS 443
Db 416 KKDPRT--ADRTIGGYEFFLLANDVDNSNPVQAEQLNWLHLYLNFSGSIYANDADANFDS 473
Qy 444 IRVDAVNVDADLLQIISDDYIKAAAYGIDKNNKNNHNSIIVEAWSNDNTPYLHDDGDMLM 503
Db 474 IRVDAVNVDADLLQIAGDIKAAKGIHKNDKAAANDHLSILEAWSNDNTPYLHDDGDNMI 533
Qy 504 NMDNKFRLSMLWSLAKPLDKSGNLPLIHNLSIVREVDREVEVTPSVSPARAHDSVQD 563
Db 534 NMDNRLRLSLYSLAKPLNQBSGMNPLITNSLVNRDNDNAETAAVPSYSFTRAHDSVQD 593
Qy 564 IIRDIKAEINPNPGYFTOEIEQAPKIYNEDLKKTKDKYTHNVNPLSYTLTLTKNGS 623
Db 594 LIRNIIRTEINPNVVGYSFTTEEIKKAEIYNKOLLATEKKYTHYNTALSYALLTKNSS 653
Qy 624 IPRVYGDMTDDGOYMAKNTVNYDAIESLLKARKMYVSGGOAMQNYQIGNEILTSVRY 683
Db 654 VPRVYGDMTDDGOYMAHKTINYEAIETLLKARIKYVSGGOAMRNQVGNSEIITSVRY 713
Qy 684 KGKALQSKDGDAITRTSGVGVWGNQPNFSLDGK-VVALNMGAAHANOEVRLMVSFKD 742
Db 714 KGKALKATDTGDRTRTSGVAIEGNNPSRLKASDRVVVNGAAHKNQAYRPLLLTTDN 773
Qy 743 GVATYATADAKSKAGLVKRTDENGVLFLNDDLKGANPQVSGFLQVWVPVGAADDQDIR 802
Db 774 GIKAYHSQDEA--AGLVRYTNDRGELIFTAADIKGYANPQVSGYLGWVPVGAADDQDIR 831
Qy 803 VAASDTASTDGKSLHQDAMDSRVNFEFGSNFQSPATKEEBEYTNVVIANNVDKVFWSGIT 862
Db 832 VAASAPSTDGKSHVONAAALDSRVNFEFGSNFQSPATKEEBEYTNVVIANNVDKVFWSGIT 891
Qy 863 DPEMAPQVYSTDGOFLDSVTLQNGYAFTRDVLGNSKANKYGTADOLVKAIKALHAKGLK 922
Db 892 DPEMAPQVYSTDGSFLDSVTLQNGYAFTRDVLGNSKANKYGTADOLVKAIKALHAKGLK 951
Qy 923 VNADWVPDQMTYFPKQEVVTVTRTDKFGKTAGSQINHSLVYVTDTKSSGGDDYQAKYGGAF 982
Db 952 VNADWVPDQMTALPEKEVTVATRVDKYGTVPAGSQIKNTLVYVDGKSGKQQAKYGGAF 1011
Qy 983 LDELKEKYPELFTKQISTGQAIDPSVKIKQWSAKYFNGSNILRGADYVLSQDVSNKYF 1042
Db 1012 LEELQAKYPELFAKQISTGVPMDDPSVKIKQWSAKYFNGTNILRGAGYVLKDAQNTYF 1071
Qy 1043 NVASDTLFLPSLLG----KVVEGIRYDGKGYIYNSSATGDQVKASFITEAGNLYYFG 1097
Db 1072 SLVSDNTFLPKSLVNPNNHGTSSSVTGLVFDGKGYVYVYST-SGNQAKNAFISLGNWNYVFD 1130
Qy 1098 KDGVMTYCAQTINGANYFFLENGTALRNTIYTDAGNSHYVANDGKRYENGQYQFGNDWR 1157
Db 1131 NGVMTYCAQTINGANYFFLENGIQLRNTIYDNGKVLVSYNDGRRYENGYYLFGQWR 1190
Qy 1158 YFKDGNMAVGLTVDNGVYDFKDGVAQAKIIVTRDGKRVYFDQHNGNAAATNTFIADKT 1217
Db 1191 YFQNGIMAVGLTRVHGAVOYFDASGFQAKGQFITTADGKLYRFDSDSGNQSINRVRNSK 1250
Qy 1218 GHVYVLGKGVAVTCAQTVGKLYFEANGQVKGDFVTSDEGKLYFYDVDVSGDMWTTF 1277
Db 1251 GEWFLFDHNGVAVTGTVTTFNGORLYFKPFGVQAKGEFIRDANGVLYRYDYPNSGNEVRNF 1310
Qy 1278 IEDKAGNWFYLGKGAAYTGAQTIRG 1303
Db 1311 VRNSKGWFLFDHNGIATGARVNG 1336
```

RESULT 4
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C.Species: Streptococcus mutans

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C/Accession: A45866
R/Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A/Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A/Reference number: A45866; MUID:91100958; PMID:2148600
A/Accession: A45866
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1431 <HON>
A/Cross-references: UNIPARC:U0100017AC5C; GB:M29296
C/Keywords: glycosyltransferase; hexosyltransferase
F/181-201/Domain: cpl repeat homology <CP1>
F/1127-1146/Domain: cpl repeat homology <CP2>
F/1192-1211/Domain: cpl repeat homology <CP3>
F/1257-1276/Domain: cpl repeat homology <CP4>
F/1277-1297/Domain: cpl repeat homology <CP5>
F/1321-1340/Domain: cpl repeat homology <CP8>
F/1341-1361/Domain: cpl repeat homology <CP6>
F/1385-1404/Domain: cpl repeat homology <CP7>

```
Query Match 43.1%; Score 3599; DB 2; Length 1431;  
Best Local Similarity 48.4%; Pred. No. 1.7e-161;  
Matches 724; Conservative 241; Mismatches 414; Indels 118; Gaps 26;  
Qy 1 MEKVRPMHKVKRWVTLVSASA--TWLASALGASVAS-----ADTDTS 44  
Db 1 METKRYRMHKVKRWVTVAVASGLITLGTTLGSSVSASVTEQQTSDKVVTKSDEDDKAA 60  
Qy 45 DDSNQA--VVTGDTTNNQATDTSIAATATS--EQSASTDAATDQASAAEQTQCTTAST 100  
Db 61 SSSQTDAPKTKQAQTQTQTAQSOANVADTSTSIKTPSQNIYTTQANSDDKTVNTKSE 120  
Qy 101 D--TAAQTNTNANBAKVPPTENENQGTDEMLAEAKNVAT--AESDIPSIDLAKMSNVKQV 157  
Db 121 EAQTSSEERTKQSEBAQ--TTASSQALTQAKAEILTQKQTAAQENKGNPVDLAAIPNVKQI 177  
Qy 158 DGKYYVYDQDGNVKNFAVSGDKIYYFDE--TGAYKDTSKVDADKSSSAVSQNTAIFAAN 216  
Db 178 DGKYYVYDGSQPKKFNALTVNNKVLFPDKNTGALTDTSTQYQFKQGLTKLND---YTPH 234  
Qy 217 NRAYSTSAKNFPEAVDNYLTADSWYRPKSLKDGKTTWETSGKDDPRPLLMWMPDTEKRN 276  
Db 235 NQIYNFNTSLETIDNVTADSWYRPRDKLKNKTKWTASSSEDRLPLMSWHPDKQTQIA 294  
Qy 277 YVNTMNVK-VGIDKITYTAETSQADLTAAAEILVQARIQKITSENNITKWLREAIASAFVKQ 335  
Db 295 YLNTMNNQGLGTGENYTADSSQESLNLAATVQVVKIETKISQQTQQTQWLRLDIINSFVKTK 354  
Qy 336 PQWNGSEKPYD----DHLQNGALLFDNQTDLTPDTQSNYRLNRTPTNQTGSLDSRPTY 391  
Db 355 PNWNSQTESDTSAGEKHLOGGALLYSN--SDKTAYANSYDYLNRTPPTSQTG----KPKY 409  
Qy 392 NPNDPLGCGYDFLLANDVDNSNPVQAEQLNWLHLYLNFSGSIYANDADANFDSIRVADV 451  
Db 410 FEDSSGCGYDFLLANDIDNSNPVQAEQLNWLHLYLNMYSIVANDPEANFDCGVRVADV 469  
Qy 452 VDADLLQISSDYLKAAAYGIDKNNKNNHNSIIVEAWSNDNTPYLHDDGDNLNMDNKFRL 511  
Db 470 VNADLLQIASDYLKAHGVYDKSEKNAIHNLSILEAWSNDNPQYNKDTKGAQLPIDNKLRL 529  
Qy 512 SMLWSLAKPLDK-----RSLNPLIHNLSIVREVDREVEVTPSVSPARAHDSVQD 563  
Db 530 SLLYALTRPLEKASNNKNEIRSGLEPVTITNSLNRSRASEGKNSEMANIYIFTRAHDSVQ 589  
Qy 564 IIRDIKAEINPNPGYFTOEIEQAPKIYNEDLKKTKDKYTHNVNPLSYTLTLTKNGS 623  
Db 590 VIAKIIRKQINPKTDGLTFLDELKQAFKIYNEDRQAKKYYTQSNITAYALMLSNKDS 649  
Qy 624 IPRVYGDMTDDGOYMAKNTVNYDAIESLLKARKMYVSGGOAMQ-NYQIGNGE----- 676  
Db 650 ITRLYYGDVSDGOYMATKSPYYDAIDTLKARIKYAAGQDMKITVYEGDKSHMDWDY 709
```



```
QY 677 ---ILTSVRYGKALKQSDKGDATRTTSGVGVVGNQPNFSLD-GKVVALNMGAHANQBY 733
Db 710 TGVLTSVRYGTGANEATKQGMVITSNPNSLKNQNDKVIYMMGAHKNQBY 769
QY 734 RALMVSTKDGVAATADADASRAGLVKRTDENGVLFLNDDLKGVANPOVSGFLOVWVPV 793
Db 770 RELPLITTKDGLTSYSDAAK--SLYRKTNDRKGLVFDASDLOGLYNPOVSGVLAVWVPV 827
QY 794 GAADQDQIRVAASDASTDGKSLHQAAMDSTRVMEFGFSNFQSFKATKEBEYTNVVIANNV 853
Db 828 GASDNQDVRVAASNKANATQVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNV 887
QY 854 DXFVSGITDFEMAPQYVSTDGQFLDSVIONGYAFTDRYDLGMSKANKYGTADOLVKAI 913
Db 888 QLFKSGVTSFEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQQDMINAV 947
QY 914 KALHAKGLKVMADWPDDQMYTPKGEVTVTTRDKEGKPIAGSQINHSLVYVTDKSSGDD 973
Db 948 KALHKSGLQVIADWPDDQYINLPGKEVTVATRVNDYGEYRKDSEIKNTLYAAANTKSGKD 1007
QY 974 YOAKYGGAPLDELKELPELTKQISTQAIDPSPVKIKQMSAKYFNNGSNIILGRGADVYL 1033
Db 1008 YOAKYGGAPLSELAAYPSIFNRTQISNGKIDPSEKITAWAKYFNGNIIILGRGVGYL 1067
QY 1034 SDQVKNYFNVASDTLFLPSSLLGKVESGIRYDKGYIYNSSATGDOQVKASFITEA-GN 1092
Db 1068 KDNASDKYFELKGNQTYLPKQMTNKEASTGFVNDGNGMTFYST-SGYQAKNSFVQDAKN 1126
QY 1093 LYFEGKDGVMVTAQTINGANYFFLENGTALNTIYTDAGNSHYVANDKRYENGYOQF 1152
Db 1127 WYFPPNNGHMYGLQQLNGEVQYFLSNGVQLRESFLENADGSKNYFHLGNRNSNGYYSF 1186
QY 1153 GND--WRYF-KDGNMAGLTVTDGNNQVYFDKDGQVAKDKIIVTRDGKVRYPDOHNGNAAT 1209
Db 1187 DNDKWRVFDASGVMAVGLKTINGTQYFDQGYQVKGAWITGSGKKRYFDDGSGNNAV 1246
QY 1210 NTFIADKTGHMYLKGQVAVTGAQTQVKQKLYFEANGQVQKGFVTSDEGKLYFYDVDS 1269
Db 1247 NRFANDKNGDWYFLNSDGLVGVQTINGKTYFYFGDGKQIKGKIIT-DNGKLYFLANS 1305
QY 1270 GDMWTDTEBKACGNWFYLGKDGAAVTAQTIRGOKLYFKANGQVQKGDIVKGTGDKIRY 1329
Db 1306 GELARNIFATDSQNNWYFYFGSDGAVTGSQTIAGKKLYFASDQKQVKGFSVFT-YNGKVHY 1364
QY 1330 YDAKSGEVOFNKVAADGKTYVIGNDGVAVDPSVVKVQTFKDGALRFRFYNLKGQLVTG 1389
Db 1365 YHDSGELQVNRFEADKDG----- 1383
QY 1390 SGWYETANHDMWYIOS-KGALTEQTINGOHLYPKEDGHQVKGQLVGTGTGDKVRYD 1445
Db 1384 -----NWYILDSNGEALTGSRINDQRVFFTFREGKQVKGDVAYDERGLLRYD 1431

RESULT 5
A44811
Glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A44811; S22726; S28809
R:Giffard, P.M.; Simpson, C.D.; Milward, C.P.; Jacques, N.A.
J.Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GIF>
A:Cross-references: UNIPROT:Q00600; UNIPARC:UPI00000BEF31; EMBL:Z11873; NID:G47526; PIDN
A:Note: sequence extracted from NCBI backbone (NCBI:81050, NCBI:81052)
C:Genetics:
A:Gene: gtfu
C:Keywords: glycosyltransferase; hexosyltransferase
F:1307-1326/Domain: cpl repeat homology <CP4>
```

```
Query Match 39.8%; Score 3325.5; DB 2; Length 1518;
Best Local Similarity 45.0%; Pred. No. 1.4e-148;
Matches 710; Conservative 239; Mismatch 456; Indels 173; Gaps 33;

QY 1 MEKNRFPKMKVKKXWVTLSSVASATMLASALG-----ASVASAD-----TDTASDD---- 46
Db 1 MENKTHYKHLKRVKQKQWVTVIAVASVA-LATVLGGLSVTTSSVSADBTQDKTVTVQSNSGTTA 59
QY 47 ----SNOAVVTVGDQVTTNQAOTD-----QTSIAATAT-----SQBSASTDAATDAQSAA 90
Db 60 SLVTSPEATKADKKTNTKEADVLTPAKETNAVEATTTNTQATAEAAATATTADAVAA 119
QY 91 EQTGTTTASTDTAAQTNTTNAME-AKWVPTENENOGFTDEMLAEAKNVATAESDPSIDLA 149
Db 120 VPKNEAVVTTDAPAVTTEKABEQPATVKAEEVNV-----TEVKAPEALAKSEVEAAL 171
QY 150 KMSNVKQVGDGKYVYVVDQGNVKNKFAVSVGDKIYYPDETGAYKDTSKVDADKSSSAVSQN 209
Db 172 SLKNTIKNTDGKYVYVYVNEGSHKENPAITVNGQLLYFGDKGALTSSSTYSFTTCTTNIVDG 231
QY 210 ATIFAANNRAYSTSAKNEAVDNYLTADSWRPKSIKDKGKTWTSKGDDPRPLMAWMP 269
Db 232 ---FSINRAVDSSSEASFELIDGYLTADSWRPASIIKDGVTWQASTAEDFRPLMAWMP 288
QY 270 DTETKRNVTYVNMKVYVIGIDKTYTATSTQAADLTAALAEVQARIEQKITSENNTKWLREAS 329
Db 289 NVDTQVNYLNTMSKVFNLDKAYSTTDKQETLKVAAKDIQIKIEQIKIQAESKQWMLRETIS 348
QY 330 AFVKTQPOWNGESEKPY-----DHLQNGALLFDNQDTLTPDTQSNRYLLNRTPTNQTS 384
Db 349 AFVKTQPOWNGKETEN-YSGKGGEHLGGALLVYND-S-RTPWANSDYRLNRTATNQGT 406
QY 385 LDSRFTYVNPDP--LGGYVDFLLANDVDNSNPVQAEQLNWLHLYLNFSGIYANDADANFD 442
Db 407 IDKSILDSQSDPNMGFGFFLLANDVDLSNPVQAEQLNQLHLYLNMWGSIVMGDKDANFD 466
QY 443 SIRDAVDNDVADALLQISSDYLKAAAGIDKNNKNANNHVSIVVEAWSLNDNDTPYLHDDGNL 502
Db 467 GIRDAVDNDVADMLQLYNTYFREYVYGVNKEANALAHISVLEAWSLNDNDYNDKTDGAA 526
QY 503 MNMNDKFLSLMLSLAKPLDKRS-GLNPLIHNSLVDRVDOR-----EVETV 548
Db 527 LAMENKQRLALLFLSLAKPIKERTPAVSPLYNNTFTTQRTDEKTDWINKDGSKAYNEDGTV 586
QY 549 P-----SYSPARAHDSVEQDIIRDIKABINPNPSFGYSFTOEETQAFK 592
Db 597 KQSTIGKNEYKGDASGNTVFIRAHDNNVQDIIAEIHKKEINPKSDGFTITDAENKQAFE 646
QY 593 IYNEDLKKTKKTYHNVPLSYTLNLTNKGSIPIRVYVYGDGMFTDDGQYMANKTVNYDAIES 652
Db 647 IYNKMLSSDKKYTLNIPAAVAVMLQNWETITRVYVYGLYTDGHHYMETKSPYDITVN 706
QY 653 LLKARKMKTSGGQAMQNYOI-----NGEILTSVRYGKALKKQSD-KGDAT 697
Db 707 LMSRIKYVSGGQARSYWLPTDGMNDSDVELYRTNEVYTSVRYGKQIMTANDTEGSKY 766
QY 698 TRTSGVGVVGNQNFSLDQKV-VALLMGAHAHQEVALMVSTKQGVATYATADASKA 756
Db 767 SRTSGQVTVANNPKNLNDQSAKUNVENMGKTHANQYRALLIVGTADGKNFTSDADAIAA 826
QY 757 GLVXRTDBNGYLYFLNDDLKGVANPQVSGFLQVWVPVGAADDQDQIRVAASDSTASDQK-S 815
Db 827 GYVKTDSNGVLTGANDIKGYEYFDMSGFVAVVWVPVGSNDQDQIRVAPSTEAKKEGELT 886
QY 816 LHQDAAMDSTRVMEFGFSNFQSF--ATKEBEYTNVVIANNVDFKVSWSGITDFEMAPQYVSS 873
Db 887 LKATEAVDSQLIYEGFSNFQIPDPSVYVTRNRKAENVDLFKSWGVTSEMAPQFVSA 946
QY 874 TDGQFLDSVIONGYAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWPDDQMY 933
Db 947 DDGTFLDLSVIONGYAFADRYDLAMSNNKYSGKEDRLDALKHAKGIQATADWPDQIY 1006
QY 934 TFPQEVVTVTRTDKFGKPIAGSQINHSLYVYTDTKSSGDDYQAKYGGAPLDELKELKEYP 993
```



```

QY 1133 GNSHYANDGKRYENGYYQFQ---NDWRYPKQGN--MAYGLTTVDGNVQYFD-KQGVQAK 1186
DB 1242 GHVYYDPRGQVAFNGFYFAGPRQDVRYPF-DGNGQMYRGLHDMYGTTFYFDEKTIQAK 1300
QY 1187 DKLIIVRDGKRVYFDQHNGNAANTFIAD-KTGHWTYLGKGVAVTGAOTVKGQKLYFEA 1245
DB 1301 DKFIRPADGRTRIFPDITGNLAVNRFAQPNKAWYILDSNGAVTGLQTINGKQYFDN 1360
QY 1246 NGQVKGDFVTSDEGKLYFYDVSDDMTDTDFIEDKAGNWFYLGKDGAAVTAQTIRGOK 1305
DB 1361 EGRQVKGFHFTVNNQR-YFLDGSGBEAPSRFVTEN-NKYYVYDGGKLVKGAQVINGNH 1418
QY 1306 LYFKANGQVKGDIKVTGDKIRYIDAKSGEYFNKTVKAAADGKTVVIGNDGVAVDPSVV 1365
DB 1419 YFNNNDYSQVKGAWANG-----RYDGDGSGQAVSNQFI-----1451
QY 1366 KGQTFKADSGALRFYNLKGQLVGSGWYETANHDWYI-QSGKALTGEOTINGOHLYPKE 1424
DB 1452 -----QIANQWAYLNDQGHKVTGLQINNKVYFYS 1483
QY 1425 DGHQVKGQLVGTGDKRVYDANSQDQAFNKSVTYVNGKTYFNGDGTAGTGNPKGQTEK 1484
DB 1484 NGAQVKGKLLT-VQGKKCYFDAHTGEQVVRV-----1515
QY 1485 DGSDFRYSMEGQLVGSGWYENAOQWLYVKN-GKVLTLGLQTVGSGQRVYFDENGIOAKG 1543
DB 1516 -----EARGCWYTFNSAGQAVTQQVINGKQLYFDGSGRQVKG 1554
QY 1544 KAVRTSDGKIRYFDENSGSM 1563
DB 1555 RYVYVG-GKRLFCDAKTGEL 1573

RESULT 7
S22737
Glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: S22737; S28810; B44811; S22727
R:Jacques, N.
submitted to the EMBL Data Library, March 1992
A:Reference number: S22726
A:Accession: S22737
A:Molecule type: DNA
A:Residues: 1-1599 <JAC>
A:Cross-references: UNIPROT:Q00599; UNIPARC:UPI00000BEF34; EMBL:Z11872; NID:G47530; PIDN
A:Experimental source: ATCC 25975
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase genes
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: S28810
A:Molecule type: DNA
A:Residues: 1-51 <GF>
A:Cross-references: UNIPARC:UPI000017027C; EMBL:Z11873
C:Genetics:
A:Gene: gtfK
C:Keywords: glucosyltransferase; hexosyltransferase
F:1456-1475/Domain, cpl repeat homology <CPR>

Query Match 38.5%; Score 3214.5; DB 2; Length 1599;
Best Local Similarity 41.3%; Pred. No. 2.6e-143;
Matches 699; Conservative 267; Mismatches 516; Indels 211; Gaps 41;

QY 1 MEKNVRFKHKVKKWVLSVASATWLASALCASVA-----SADTDTASD 45
DB 1 MENKRYKLHKVQKWVTLAVASVA-LATIVGGSVATSLASAEETNNSGSPSTTVGE 59
QY 46 DSNQAV-----VTGD--OTTNNQATDQTSIAAT-----ATPQASASTDAATDQ 86
DB 60 NTNPFVEKEVGTTTEVANTTTERAEVTDKADGAGTTVQPNSGTTSURAAAVEAKP 119
QY 87 ASAAEQTQGTASTDTAAQTNTNANEAKWVPTENENQGFDEMLAEAKNVATAESDIPS 146

```

```

DB 120 ETTAKPEVATKPEATATTSEVAANAGVA--APTTERK-----KELSEAEIKAAVSLDNIK- 171
QY 147 DLAKSNVQKQVGGKYYVDDGNNVKNFAVSVDGIYYFDETCAYKDDTKVDADKSSAV 206
DB 172 -----KEKDGKYYLLEDGSHKKNFAITVNGQVLYFDENGALSSSTYSYSTQSTNNL 223
QY 207 SONATIFAANNRAYSTSAKNFEAVDNYLTADSWYRPFKSLKDGKWTGSKDDFRPLLMA 266
DB 224 ---VTDFTKNNAAYDSTKASFELVDGYLTADSWYRPFKEILEAGTTWKASTEKDFRPLMS 280
QY 267 WHPDTEKRYNYNNKVVG-----TDKTYTAETSOADLTAAABELVOARIEQKITSENTRYK 322
DB 281 WHPDKDTQVAYLYNMTKALSNGEETKDVFTIENSQASLNAQAQILQKIEVKAANKSTD 340
QY 323 WLREISAFAVTKPQWNGSEKPYDDHLQNGALLFDNOTDLTPDTCOSYRLNRTPTNQT 382
DB 341 WLQSQIEAFVKQDKKNINSESPGKEHFQKGALLFVN-SDSTKWANSDRKLNQATSY- 398
QY 383 GSLDSRFTYNPNDPLGGYDFLLANDVDNSNPVQAELNMLHYLLNFGSIY--ANDADAN 440
DB 399 --IKNHIKIVNGSD--GGYEFLLSNDIDNSNPVQAEMLNQLYYFNNWGOIVFEGDKDKDAH 454
QY 441 FDSIRVDVNDVADLLQLISSDYLLKAAYGIDKNNKANNHHVSIWEAWSNDTPYLLHDDGD 500
DB 455 FDGIRVDVNDVSDMLQLVSSYMKAAKYKNSSEARALANISILEAWSHNDPYVYVNEHT 514
QY 501 NLNMNDKFRLSMLSLAKPL-DKESGLNPLHNSIVD-----REVDDREVEVTVPS 550
DB 515 AALSDMDNGLRLSIVHGLTRPVTKGTGARNAMKDLINGYFGLSNRAEVTSYDQLGFAF 574
QY 551 YSFARAHDSQDIIIRDIKAEINPNSFGYSFTQBEIEQAFKIYNEDLKTDKKYTHYV 610
DB 575 YLFEVRAHDSQVTVIADIISKKIDPTDGTFTLQDKQAQFIYNADMLKVDKEYTHSNI 634
QY 611 PLSYTLLLTNKGISIPRVYGDMDTDDGQYMANVTNYDAIESLLKARMKYVSGGQAMQNY 670
DB 635 PAAYALMLQTMGAATRVYVYDLYTDNGQYMAKSPYFQITTTLLKARPKYVAGGQTSYH 694
QY 671 QIG-----NGBILTSVRYGKALQSD--KGDATTSTSGVGVVGMGNQPNFSL-DGKV 719
DB 695 NLAGDGVSSAKDNKEVLSVRYGQDLMKSTDTDEGGKYGRNSGMLTIANNPDLKADGET 754
QY 720 VALANMAAHANOEYRALMVSTKDGVTATYATADASKAGLVKRTDENGILYFLNDLKGVA 779
DB 755 ITVNGAAHKQAYEPPLLLGTEKIGVSSINDSDTK---LVKTYDAQENLVFTADEIKGPK 811
QY 780 NPQVSGFLQVWVVGAAADDQDIRVAASDTASTDGKSLH-QDAAMDSTRVMFEFSNQPSPA 838
DB 812 TVDMSGYLSVWVVPVCGATDDQNVLAKPSTKAYKEGDKVYSSSALEAQVIYEGFSNFQDFV 871
QY 839 TKEEETVNVIVANNVDFKFSWGIITDFEMAPQVYVSTDGQFLDSVQFONGVAFTRDYDLGMS 898
DB 872 KEDSQYTNKLIANAADLFSWGIITDFEMAPQVYVSTDGQFLDSVQFONGVAFTRDYDFAMS 931
QY 899 KANKYGTADQLVKAIKALHAKGLKVMADWVPDQMTFFPKQEVVTVTRTDKFGKPTAGSOI 958
DB 932 KNNKYGSKEDLRDALKALHKQIGQIVADWVPDQMTFFPKQEVVTVTRTDKFGKPTAGSOI 991
QY 959 NLSLYVTDTKSGDDYQAKYGAFDELKKEYPFLTKKQISTGQAIDPSVVKIOWSAKY 1018
DB 992 VNKLVTNTKSSGNDFOAQYGGAFDLQKLYPEIFKEVMEASGKTIDPSVVKIOWSAKY 1051
QY 1019 FNGSNILRGADYVLSDOVSNKYFNVAASDTLFLPSSLIGKV-VESGIRYDGGYLYNSA 1077
DB 1052 FNGNTIQRGSDYVLSLSD--GKLYFTVNDKGTFLPAALTGDTKAKTGFAYDGTGVYTYTT- 1108
QY 1078 TGDQVKASFITEAGNLYYFGKDGVMVTGAQTINGANYFFLENGTALRNTIYVDAQNSHY 1137
DB 1109 SGTQAKSOFVTYNGKQYVFNKGLVLTGEQITDGSNYFFLPGVNGVMTDQVRKNAQGSULV 1168
QY 1138 YANDGK-----RYENGYQFQGNDRWRYFKDGNMAVGLTVDGNGVQYFDK 1180

```

Db 1169 YKSGKLTITQCKEYVTVKDDSGKEKFQYF-----FKGGIMATGLTEVEGKEKYFD 1222
Qy 1181 DGVQAKDIIVTRDQKRVYFDQHNGNAATNTFIADKTHGHWYLGKDGVAVTAQTVGKQK 1240
Db 1223 NGYQAKGVFVPTKDHLPFCGDSGERKYSGFF-EQDGNWYVANDKGYVATGFTKVGQN 1281
Qy 1241 LYFEANGQVKGDFVTSDBGLYFYDVDSGDMWTDTFIEDKAGNFWFLKDGAAVTAQ 1300
Db 1282 LYFNEKGVOVKNR-----FQVGDA-----TYANNEGDLVRGAQ 1317
Qy 1301 IRGQKLYFKANGQQKGDIVKGTGDKIRYDADKSGEQVFNKTVKAADGKTVYVIGNDGVAV 1360
Db 1318 INGDELYFDESQKQKGFVNNPDGTTSYDAITGVK-----V 1356
Qy 1361 DPS-VVKGQTFK-DASGALRFYNLKGQVTS---GWYETANHFWYIQS-GKALTGSGT 1414
Db 1357 DTSLVVDGQTFNVDAKG-----VVTKAHTGFTYTTGNNWFYADSYGRNVTAQV 1406
Qy 1415 INGOHLYFEKDHQVKGQVTCGTGKRVYDANSQDAFNKSVTVNGKTYFEGND-GTAQ 1473
Db 1407 INGOHLYFDANGRQVKGGFVNTDGRSFYHWTGDKLVSTFPATGHRWYVYADRGNVV 1466
Qy 1474 TAGNEPKGQTFKDGSDIRFYSMEGQVLTG-----SCWYENAOQ 1511
Db 1467 T-----GAQVINGQKL-FPDTGQVKGAFATNANGSRSYHWTGKNKLVSTFPFGSDNN 1520
Qy 1512 WLYV-KNGKVLTLQTVGQVRVYFENGIOAKGKAVRTSDGKIRYFDENSGSMITNQW-K 1569
Db 1521 WYADAKGEVVVEQTINGOHLFYDQTKQVKGATATNPDGSGSYDVHTGKAINRWK 1580
Qy 1570 FYVQYVYFEGNDG 1582
Db 1581 IPSGQWVYFNAQG 1593

RESULT 8
T30857
glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30857
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infected. Immun. 63, 409-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: Z20909; PMID:95122197; PMID:7822030
A:Accession: T30857
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166E; EMBL:L35495; NID:g662378; PID
C:Genetics:
A:Gene: gtfI

Query Match 37.5%; Score 3128.5; DB 2; Length 1449;
Best Local Similarity 42.9%; Pred. No. 2.5e-139; Mismatches 231; Indels 199; Gaps 31;
Matches 670; Conservative 231; Mismatches 460; Indels 199; Gaps 31;
Qy 1 MEKNRFRKHKYKKGKRWVLSVASATM---LASALG-----ASVA---SADT 40
Db 1 MDKVKHYKHKYKKGKRWVLSVASATM---LASALG-----ASVA---SADT 60
Qy 41 DTASDDSN-QAVVTGQDTNNQATDQTSIAAT----- 71
Db 61 GTVSNDTTAAQDPPTTAAATNDVATDOATPTATDTLTTTNTVAANA VDTVATVGTDR 120
Qy 72 -ATSSQASTDAATQAAAEOTQGTASTDTAAOITTTNANEAKWVPTENENQGTDEM 130
Db 121 ATTNDITATNDTAVDTTN-NNTTDTTVDTRAATERRATGARRGPTGRRATPVNGT 179
Qy 131 AEAKNVATAESDIPSDLAKM-----SNVKQVDGKTYYYDQDGNKKNFAVS-VGDKIYF 185
Db 180 NNANNTVTVNNDLPATNNVTDGFSHIKTINGKQYVVEDDGTIRKNVYLERIGGSQYFN 239

Qy 186 DETGAYKDTSKVDADK-----SSSAVSQONATIFAANNRAYSTSAKNFEADVNNYLTADSW 239
Db 240 AETGELSNOKEYRFDKNGGTGSSADSTNTNTVNGDKNAFYGTDDKQLELVDGYETANTW 299
Qy 240 YRPKSIKLDGKTWTESGKDDFRPLLMAMWPDTEKRNVTNVMNKV-VGIDKITYTAETSOA 298
Db 300 YRPKEILLKDGKEWTASTENDKRPLLTVWMPKAIQASLYLNMKEGGLGNTQTYTSFSQT 359
Qy 299 DLTAARAEVLQARIQKITSSENTKWLREAI SAFVKVTPQWNGESEK-PYDHLQNGALLF 357
Db 360 QMDQAALAEVQKRIERIAREGNTDMLRTTKNFVKVTPQWNGSTSENLDNDHLQGLL 419
Qy 358 DNQTDLPDPTOSYRLNRTPNQTGSLDSFTYNPNDPLGGYDFLLANDVNSNPVQA 417
Db 420 NND5-RTSHANSYRLNRTPNQTGSLDSFTYNPNDPLGGYDFLLANDVNSNPVQA 476
Qy 418 EQLNWLHYLLNFGSIYANDADANFDSIRVDADVNDVADLLQISSDYLKAAYGIDKKNKA 477
Db 477 EQLNWLHYIMNIGITGSGEDENFDGVRDAVDNVDNADLLQIASDYFKAYGADQSOQA 536
Qy 478 NNHVSIVGAWSDNTPYLLHDDGNLMMNDKFRSLMWSLAKPLDKRGLNPLIHLNSVD 537
Db 537 IKHL5ILEAWSHNDAYNEDTKGAQLPMDPMHLALVYSLLRPIGNRSGVEPLISNSLD 596
Qy 538 REVDDREVEVTPSYSPARAHDSVODIIRDIKAEINENSEFSGYSFTQEEI5QAFKYNED 597
Db 597 RSESGKSKRWNTAFVRAHDSVQSIIGQIKNEINPQSTGNTFTLDEMKAFAFIYND 656
Qy 598 LKKTDKKYTHVNPVLSYLLLTNKGSI PRVYVGDWFTDDGQVMANKTVNYDAIESLLKAR 657
Db 657 MRSANKQYQYNI PSAYALMLTHKDTVPRVYVGDWFTDDGQVMANKTVNYDAIESLLKAR 716
Qy 658 MKYVGGQAMQNYQIENG-----ILTSVRYKGKALKOSDKGATATRTSGVGVNMG 709
Db 717 IRYAAGGDMKVYIGYNTNGWDAAGVLT5VRYGTGANSASDTGTAETRNQGMVIVSN 776
Qy 710 QPNFSLDGKVALNMGAAHANOERYALMVSTKDGVAITYATDADAKAGLVKRTDENGVL 769
Db 777 QPALRLTSN-LTINNGAHRNQYRPLLLTNDGVAITYLNDSDAN--GIVKYTDGNGMLT 833
Qy 770 FLNDLKGANPQVSGFLQVWPVGAADDDIRVAASDTASTDGKSLHQDAMD5SRVME 829
Db 834 F5ANEIRGIRNPQDGYLVAVVPVGA5ENQDVRVAPSKKSSGLVY5SNAALDSQVIYE 893
Qy 830 GFSNFPQSPATKEEYNTVNIANNVDKFSVSGITDFEMAPQVYVSTDDGFLDSVIONGVAF 889
Db 894 GFSNFPQSPQVNTKKIAENANLFK5WGIT5FEFAPQVYVSSDDG5FLDSVIONGVAF 953
Qy 890 TDRYDLGSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTFPKQEVVTVTRDKF 949
Db 954 TDRYDLGSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTFPKQEVVTVTRDKF 1013
Qy 950 GKPIAGSOINSHLYVTDTKSSGDDYQAKYGAFDELDELKEKYP5ELFKKQISTGQADPSV 1009
Db 1014 GETDKGAILD5LAAKTRTFGNDYQGYGAFDELDELK5LPQIFDRVQISTGKEMTDE 1073
Qy 1010 KIKQWSAKYFN5G5NILGRGADYVLSQVSNKYFNVASDTLFLPS5LLGK5V5G5IRYDGK 1069
Db 1074 KITQWSAKYMG5N5ILDRG5EVLK5NGL-NGY5GTNG5K5V5LPK5V5G5N5Q5TNG5N5Q5D 1132
Qy 1070 G-----YIYNSSATGDQV5K5P5TE-AGNLY5FGK5G5VM5TGAQ5TINGAN5VF 1116
Db 1133 G5G5K5E5K5L5F5VR5RYNN---G5YAKNA5FI5K5ND5GN5V5Y5F5N5G5MA5VEK5TID5G5Q5FF 1189
Qy 1117 LENG5TAL5NTI5Y5TDAQ5GN5SHY5YAND5G5K5RY5NG5Q5F5G5ND5W5RY5F5K5GN5MA5V5GL5T5T5VD5GN5VQ 1176
Db 1190 LANG5QL5R-----D5YRQ-----N5RR5Q5V5F 1209
Qy 1177 YFDK5DG5Q5AK5DKI5IVTRD5G5K5RV5FDQ5HNG5NAA-TNT5FIAD5K5TH5WY5LG5K5G5VA5VTG5AQ5T 1235
Db 1210 Y5DQ5NGV-----LNANG5K5Q5PK5PD5NN5N5AS5GR5NQ5F5Q5IGN5V5W5AY5D5G5N5K5RV5T5G5QN 1262
Qy 1236 VGK5K5L5FEANG5Q5VK5G5DF5VT5D5EG5K5LY5FY5DV5D5G5DM5W5T5D5F5IED5K5AG5N5FW5LG5K5GA5V 1295

Db 1263 INQELFFDNNGVQVKGRTV--NENGAIYYDANSGEWARRFAEIBFGWYAFNNDGTAV 1321
QY 1296 TGAQTIRGQKLYFKANGQVKGDIIVKGTGDKIRIYDAKSGEQVFNKTVRAADGKTVVIGN 1355
Db 1322 KGSQNINGQDLFDQNGRQVKGALA--NVDGNLRYVDVNSGELYRN----- 1365
QY 1356 DGVAVDPSVVGQTFKADAGALRPFYNLKGOLVTGSGWYETANHWDWYIY--SGKALTGEQT 1414
Db 1366 -----RFHEIDG-----SWYFPGNGNAVKGWYN 1389
QY 1415 INGOHLYFKEDGHQVKGQVLTGVDGKVRYYDANSQDAFNKSVTVN--GKTYIFGNDGTAQ 1473
Db 1390 INQNLFFDNNGKQIKGHLVR--VNGVVRYPDPNSGEMAVNRWVEVSPGWWVYFDGSGRGQ 1448

RESULT 9
T30552
glucosyltransferase N - Streptococcus salivarius (fragment)
C:Species: Streptococcus salivarius
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30552
R:Jaffe, R.I.
submitted to the EMBL Data Library, February 1998
A:Description: Streptococcus salivarius V1477 gcfN.
A:Reference number: Z20854
A:Accession: T30552
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-1449 <JAF>
A:Cross-references: UNIPROT:O68542; UNIPARC:UPI00000810FD; EMBL:AF049609; NID:G2935545;
C:Genetics:
A:Gene: gtfN

Query Match 37.3%; Score 3114.5; DB 2; Length 1449;
Best Local Similarity 43.0%; Pred. No. 1.le-138; Indels 197; Gaps 32;
Matches 671; Conservative 227; Mismatches 464;

QY 1 MEKNVRFKMHKVKRWVTLVSASATM---LASALG-----ASVA-----SADT 40
Db 1 MDKVKHYKXHKVKKQWVTVIAVGLSLGAVSAVSLGTNDGVQVQADEHTDAVLPDITVDT 60
QY 41 DTASDDSN--QAVVTGQDTNNQATDQTSIAA-----TATSQSASTDAATQASAA 90
Db 61 GTVSNDDTTAAQDPTTAVAAATNDVATDQATPTATFDLTDTTINTVAANA VDTVATVGTDR 120
QY 91 EOTQGTASTDTAAOTTTN-----ANEAKWPTENENOGFTDEMLA 131
Db 121 ATTNDTTATNDTAVDTTNNNTTDTTNTNRAATTERRATGARRGPTGGRATPVNGNTN 180
QY 132 EAKNVATAESDSIPDLAKM-----SNVKQVQDKYIYDODGNVKKNFVAVS--VGDKIYYFD 186
Db 181 NANNVTVVNNDLPATNVVTDGPHSHIKTINGQYVVEDDGTIRKNVYLERIGSGQYFNA 240
QY 187 FTGAYKDTSKVDADK-----SSSAVSQVATIFAANNRAYSTSAKNFEAVDNYLTADSWY 240
Db 241 ETGELSNQKEYRFDKNGGTGSSADSTNTNTVNGDKNAPYGTDDKDIELVDCGYFTANTWY 300
QY 241 RPKSTLKDGKWTESGKODFRLLMAWMPDTEKKNYNNYMKV--VGIDKTYTATTSQAD 299
Db 301 RPKELTKDGKEWTASTDKRPLLVWMPFSKAIQASLYNMYKKEQGLGTNTQTYTTSFSSQTQ 360
QY 300 LTAAAEALVQARTEQKITSENNTKWLREAIASFVKTPQWNGSEK--PYDDHLQNGALLFD 358
Db 361 MDQAALEVQKRIEGRIAREGNTDWLRTYIKNFVKTPQGNSTSENLDNNDHLQGGALLYN 420
QY 359 NOTDTPPTQSNYRLNLTPTNTQTSGLDSRFTPNPNDPLGGYDFLLANDVNSNFPVQAE 418
Db 421 NDS--RTSHANSYRLLNLTPTTSQTKHNPYTKDTSN--GGPEFLANDIDNSNFAVQAE 477
QY 419 QLNWLHYLLNFGSIYANDADANFDSIRVDAVDNVDADLLQTSISDYLKAAYGIDKKNKAN 478
Db 478 QLNWLHYIMNITGITGSGEDEFDGVVRVDAVDNVDNADLLQIASDYFKAKYGADQSDQAI 537

RESULT 10
T31098
probable dextranase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides

QY 479 NHVSIIVEAKSNDTPYLDHDDGNLMMNDKFRLSMLWSLAKPLDKESGLNPLIHNSLVDR 538
Db 538 KHLISILEANSHDAYNEDTKGAQLPMDDPHMLAVYSLLRPIGNRSGVEPLISLNSLDR 597
QY 539 EVDREVEVTPSYSPARAHDSVQDIIRDIKAEINPNRSGYSGFTQEEIEQAFKIYNEDL 598
Db 598 SESGKNSKEMANYAFVRAHDSVQSIIGQIKNEINPQSTGNTFTLDEMKAPEIYNKOM 657
QY 599 KXTDKKYTHYNPLS VTLTLTKNGSIPRVYVYGDMTDDGQYMANKTVNYDAISLAKARM 658
Db 658 RSANKQYTOYINPSAYALMLTHKDTVPVRYGDMYTDGQYMAQKSPYDAIETLLKGR 717
QY 659 KTVSGQAQMYQIENG-----ILTSVRYKGGALKQSKDKGATTTTSGVGVVMGNQ 710
Db 718 RYAAAGQDMKVNYIGVNTNGWDAAGVLTSVRYGTGANSADTGTAETRNQGNMAVIVSNQ 777
QY 711 PNFSLDGKVALNMGAAHANQYRALMVSTKQGVATYATDADAKAGLVKRTDENGYYLF 770
Db 778 PALRLTSN--LTINMGAAHRNQYRPLLLTNTDGVATYLNDSAN--GIVKYTDCGNLTF 834
QY 771 LNDLKGVANPOVSGFLOVWVPVGAADODIRVAASDTASTDCKSLHQDAAMDSDRMFEG 830
Db 835 SANEIRGINRPQVVGYLAVWVPVGAENQDVRVAPSKENKSSGLVVTESNAALDSQVIYEG 894
QY 831 FSNFQSFATKEEYTNVVIANNVDFVSMGIITDFEMAPQYVSVSTDCQFLDSVIQNGYAF 890
Db 895 FSNFQDFVPNPQSYTNKKIAENANLFSKGIITSEFAPQYVSSDDGSLDSVIQNGYAF 954
QY 891 DRYDGLMSKANKYGTADQLVKAIKALHAKGLKMAADWPDQMYTFPKQEVVTVTRDKFG 950
Db 955 DRYDGLMSKANKYGSGLADLKAALKSLHAGVISAIDWVPDQYVNLPGDGVATVTRVNNY 1014
QY 951 KPIAGSQNHSLYVTDTKSSGDDYQAKYGGALDELKEYPELFTKQISTGOAIDPSVK 1010
Db 1015 ETKGAIIDHSLYAAKTRTFGNDYQGYGGAFLDELKRLYPOIFORVQISTGKMTTDEK 1074
QY 1011 IKQSAKTFNGSNIILRGADYVLSQVSNKYFNVAASDTLFLPSSLLGKVVESGIRYDGGK 1070
Db 1075 ITKMSAKYNGTNIILDRGSEYVLKGL--NGYGTNGGKVS LPKVGSNSTGNDGNQGDG 1133
QY 1071 -----YIYNSATGQVKAFFITE--AGNLYYFGKDGVMVTGAQTINGANYFFL 1117
Db 1134 SGKFEKRLFSVRYRYNN---GOYAKNAFTKNDGNVYFPDNGRMAVGEKTIIDGQYFFL 1190
QY 1118 ENGTALRNTIYDAGNSHYANDCKRYENGQYQOGNDWRYFKDGNMAVGLTIVDGNVQY 1177
Db 1191 ANGVQLR-----DGYRQ-----NRRGQVY 1210
QY 1178 FDKGV--QAKDKIIIVTRDGKVRYPQHNAGNAATNTFIADKTHWYVYLGKGVAVTGAQTV 1236
Db 1211 YDQNGVLSANGK----QDPKP---DNNNTSGRNFQVQIGNVWAYYDNGKRVIGHQVI 1263
QY 1237 GKQKLYFPANGQVKGDFVTSDEGLYPYDVSQDMWTDFTIEDKAGNMFYLGKDGAAVT 1296
Db 1264 NQOELFFDNNGVQVKGRTV--NENGAIYYDANSGEWARRFAEIBFGWYAFNNDGTAQ 1322
QY 1297 GAOITRGQKLYFKANGQVKGDIIVKGTGDKIRIYDAKSGEQVFNKTVRAADGKTVVIGN 1356
Db 1323 GSONINGQDLFDQNGRQVKGALA--NVDGNLRYVDVNSGELYRN----- 1365
QY 1357 GVAVDPSVVGQTFKADAGALRPFYNLKGOLVTGSGWYETANHWDWYIY--SGKALTGEQT 1415
Db 1366 -----RFHEIDG-----SWYFPGNGNAVKGWYN 1390
QY 1416 NQOHLFYKEDGHQVKGQVLTGTDGKVRYYDANSQDAFNKSVTVN--GKTYIFGNDGTAQ 1473
Db 1391 NQNLFFDNNGKQIKGHLVR--VNGVVRYPDPNSGEMAVNRWVEVSPGWWVYFDGSGRGQ 1448

C:Species: Leuconostoc mesenteroides	
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004	
C:Accession: T31098	
R:Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.	
FEMS Microbiol. Lett. 159, 307-315, 1998	
A:Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DS	
A:Reference number: Z20981; MUID:98164374; PMID:9503626	
A:Accession: T31098	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-1508 <MON>	
A:Cross-references: UNIPROT:O52224; UNIPARC:UPI00000BB69B; EMBL:AF030129; NID:g2766611;	
A:Experimental source: strain NRRL B-1299	
C:Genetics:	
A:Gene: dsrB	
C:Function:	
A:Description: produces dextran composed only of alpha(1-6) glucosidic bonds	
C:Keywords: glycosyltransferase; hexosyltransferase	
Query Match	37.0%; Score 3088; DB 2; Length 1508;
Best Local Similarity	42.3%; Pred. No. 2.1e-137;
Matches 674; Conservative 231; Mismatches 466; Indels 224; Gaps 35;	
QY	2 EKNVRKHKVKRVTLSVASATWLAS--ALGASVASADTDTASDDSNQAVVTGDTTN 59
DB	6 ERNRKKLYKSGSKSWIGLLIILSTIMLSWTATSQNNADSTNTVTDKS--VTVSNNSNT 63
QY	60 NO-----ATDQTS--IAATAT--SEQASTDAATDQASAAEQTQGT----- 97
DB	64 NQHDVTVDKQTPVKNDQTTQIAANATQAEKVASDTTDTQKQAEATNNTKEDSIDN 123
QY	98 -----ASTDTAAQT-----TNAME-AKWVPTENENOGFTDEMLAEAK----- 134
DB	124 LTKQLPAVTPPTANQTKYLEKDGKRYVYTSNDTLAKGLTTVDNHRQYFDNNGVQAKGFV 183
QY	135 -----NVAATESDSIPDLAKMSNVKQV-----DGKYYVYDQD----- 167
DB	184 TUNSKTYLLDPNSGNAVGTGIIQIGSQTLAFNDNGEQVADFPTAPDGKTYTFDDKGQATI 243
QY	168 -----GNVKCNFAVSGDKIYYFDETGAJKOTSKVDADKSSAVSQNATI 212
DB	244 GLKATNGHNYFDSLGLQKKGFTGVIDGQVRYFDQESG-QEVSITTDTSQIKEGLTSQ-TTD 301
QY	213 FAANRAVSTSKNFEADVNTLTADSWTRPKSILKDGKWTBESGKODFRPLLMAWPDTE 272
DB	302 YTAHNAVHSTDADFDNFNGYLTASWYRPKDVRNGQHWEATTANDFRPIVSVWVPSKQ 361
QY	273 TKRNVYNNKVVIGID--KTVTAETSOADLTAAAEVLQARIEBKITSSENTKWLREAI SA 330
DB	362 TQVNYLNTMSQGLIDNRQMFLSKDNQAMLNACTTVOQAIETKIGVANSTAWLATAIDD 421
QY	331 FVKTFQWNGSEKPYDHLQNGALLFDNQDTLTPDTQSNVRLNRTPTNQTSGLDSRFT 390
DB	422 FRTQPMWMSSEDPKNDHLQNGALTFTVN-SPLTPDNTSNFELLNRTPTNQGVF---K 476
QY	391 YNPNDPLGGYDPLLANDVNSNPVQAEQLNWLHLNFGSIIYANDADANFDSIRDAVD 450
DB	477 YTIDQSGKGFELLANDVNSNPVQSEQLNWLHLNFGSITANDSAANFDGIRDAVD 536
QY	451 NYDADLLQISSDYKAAVGDKNKANNHVSIVAEWSNDNTPYLHDDGDNLMNMDNKR 510
DB	537 NVDAQLQLAAYDFKAAVGVDKNDATANQHLSEIDWNSHNDPEYVKDFGNQLTWDYWH 596
QY	511 LSLMLSLAKPLDKRGLNPLIHNSLVREVDREVEYTVPSYFARAHSEVQDIIRDIK 570
DB	597 TQLIWSLTKDMRMGTQRMFDYVILVNRNHDSTENTAIPNTYSFVRAHSEVQTVIAQII- 655
QY	571 ABINP---NSFGYSTQBEIEQAFKIYNEDLKTDKKTHYVPLSYTLLLTNKSISPRV 627
DB	656 SELHPDVKNSL--APTAQQLAEAFKIYNNDEKQADKKYTYQNMPSAYAMLLTNKDTVPRV 713
QY	628 YGDMFTDDGOYMANKTVDYDAIESLLKARMKYVSGQAMQNYQIENGEBILTSVRYGKA 687

RESULT 11

A41483

glucosyltransferase (EC 2.4.1.1-) gtfS precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Mar-2004

C:Accession: A41483

R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.

Infect. Immun. 58, 2452-2458, 1990

A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltrans

A:Reference number: A41483; MUID:90316665; PMID:2142479

A:Molecule type: DNA

A:Residues: 1-1365 <GIL>

A:Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:

C:Genetics:	
A:Gene: gtf5	
C:Keywords: glycosyltransferase; hexosyltransferase	
Query Match 36.3%; Score 3031; DB 2; Length 1365;	
Best Local Similarity 41.6%; Pred. No. 8.9e-135;	
Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;	
QY	1 MEKNVRFKHVKRVRVTLISVASATMLASALGASVASADTDTASDSDQA--VVTGD--- 55
DB	1 MEKNLRYLKHKVQKQWALGVITVT--LSFLAGQVVAAADTNNDGTSVQVNRKVPDPKF 59
QY	56 ---QTNNQATDQTSIAA-----TATSEQASASTDAATDQ--ASAEQQTGTTATDTPAQT 106
DB	60 DAQAQNGQLAQAFMAAQAADQATATSQVSPATDGRVDNQVTPAANQPAANQADV--- 115
QY	107 TTNAEAKVPTENENQGTDEMLAEAKNVATAESD--SIPSDLAKMSNVQVDGKYYY 164
DB	116 ---ANPA-----TDAGALNROSAADTSDGKAVPQTSDPQGHLETVDGKTYV 160
QY	165 DODGNVKKNFVAVSGDKIYYFD-ETGAYKDTSKVDADKSSSAVSQNA-TIPAANNRAYST 222
DB	161 DANGQLKNYSVIDKTYFDQGTG---EAQTDLPTKQANQDNVPSYQANQATSN 216
QY	223 SAKNFEADVNYLTADSWTRPKSILKDKTWTSGKDDFRPLMAWMPDTETKRYNYVMN 282
DB	217 EASSPETVDNYLTADSWTRPKSILKDKTWTSGKDDFRPLMAWMPDTETKRYNYVMN 276
QY	283 KVVGDIKTYTATSOADLTAAELVQARIEQKITSENNTKWLREALSAFVKTPQWNGES 342
DB	277 KEGLISGSY--RONGANLDAATQNTQSAIEKKIASEGNTNWLDRKMSQFVKSQNWSTAS 334
QY	343 EK---PYDDHLQNGALLFDNQTDLPTDQSNVRLNTRPTMTQTSGLDSRETYNENDPLG 398
DB	335 ENETYPNQDNQWQGALLFNSKD--TEHANSWRLLNRNPTFQTK-QKYFTTN----YA 388
QY	399 GYDFLLANDVNSNPVQABQLNLWLYLNFSGSIYANDADANFDSIRDAVDNVDADLLQ 458
DB	389 GYELLANDVNSNPVQABQLNLWLYLNFSGSIYANDADANFDSIRDAVDNVDADLLQ 448
QY	459 ISSDYLKAAAGIDKKNKNNHVSIVEAWSNDTPYLHDDGNLNMNKKFLSLMLWSLA 518
DB	449 IORDYKAKYGTQDQEKNAIDHLSLEAWSGNDNDYVKDQNNFSLSIDNDQSRGMLKAFG 508
QY	519 KPLDKRSGLNPLHNSLDVREDDVETVPQSYSPARAHDSQVDIIRDIKAEI-NPNS 577
DB	509 YASAVRGNLSNLATAGLNRSNP--DSDPVPNVYFIRAHDSVQTRIAKILIREKGTNA 567
QY	578 FGYSFTQEEIEQAFKINEDLKTKDKYTHYNVPLSYTLTLTNKGSIPRVYVYGMFTDD 636
DB	568 DGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNPLPMAVAMLQNKDVTVRVYVYGMFTDN 627
QY	637 GOYMANKTVNTDAISLKLARKMYVSGGQAMQYQINGEILTSVRYKQKALKQSDKGA 696
DB	628 GOYMATKTPFYNAIETLLKGRIKYVAGGQAVSYKQDWSGILTSVRYKQKANSASDAGNT 687
QY	697 TTRTSGVGVWGNQNFSLDGVKVALNMGAAHNOEYRALVYTKDGVATYATDADASKA 756
DB	688 ETRNGSMALLNNRNFRA--YNNLTNMGAAHKSQAYRPLLTSTKDGATYLNDSVDV-S 745
QY	757 GLVKRTDENGLYLFYNDLKLGVANPQVSGFLQVVPVGAADDDQDIRVAASDTASTDGKSL 816
DB	746 ROYKYTDSQGNLSFSAELQSVANNAQVSMIQVVPVGAADNQDVRTSPSTQATKQNIY 805
QY	817 HODAMDSRVMPGFSNFQSFATKEEYTNVVIANNVDKFSVSGITDFEMAPQYVSSDTG 876
DB	806 HQSDALDSQVIYEGFSNFQAFQSPQYTNNAVIAKNGDLFKSWGITQFEMAPQYVSSSDG 865
QY	877 QFLDSVIONGYAFTRDYLDGWSKANKYGTADQLVKAIKALHAKGKVMWADVPDQWYTFP 936
DB	866 TFLDSVILNGYAFSDRYDLAWSKNNKYGSKODLANAIKGLSAGIKVLSDLPVNLQNLN 925
QY	937 KQEVVTVTRTKDFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKEKYPELFTK 996
DB	926 GKEVTVATRVNQYQAKSGATINKTPYVANTRSYG-DYQEQYGGKFLDDLQKLYPRLFS 984
QY	997 KOISTGQADIDPSVKIKQMSAKYFNGSNILGRGADVYLSDOVSNKYFNVAADTLFLPSSLL 1056
DB	985 KOISTGKPIDPSVKITNMSAKYFNGSNILGRGAKYVLSB--GNKYLNLTADGKFLPVLN 1042
QY	1057 GKVSFSGIRYDG---KGIYNSSATGDQVKASFITEAGNLNYFFGKDGVMVMTGAQTINGA 1112
DB	1043 NTYGQPOVSANGFISKNGGIHYLDKXNQEVKRFKEISGSWYFFSDGKMATGKTGLIND 1102
QY	1113 NYFFLENGTALNTIYTDAGNSHYANDKRYEN-GYQOF---GND-WRYFK-DGNMAV 1166
DB	1103 TYLFMPNGKQLKEGVWYDGK-KAYYYDNGRTWTNKGFEVFRVDGDKWRYENGDTGTAI 1161
QY	1167 GLTTVDGNGVQYFDKGVQAKDKIIIVTRDGKRYVFOHNGNAATNTFIADKTHWYVLGKD 1226
DB	1162 GLVSLD----- 1167
QY	1227 GVAVTGAQTVGKQKLYFEANGQOVKGDFTVSDGKLYFYDSDGMDMTDTFTEDKAGNWF 1286
DB	1168 -----NRTLYFDAYGYQVKGQTVTIN-GKSYTFDADQGD----- 1201
QY	1287 YLGKDGAAVTGAQTIIRGQKLYFKANGQQVKGDIIVKTDGKIRYDAKSGEQVFNKTVKAA 1346
DB	1202 ----- 1201
QY	1347 DGTYYVIGNDGVAVDPVSVVKGQTFKXDGALRPFYNLKGQVLTGSGHYETANHWDVYIQSG 1406
DB	1202 -----VOTDNANPAP-----QGO-----AGMKLLGDNQMGYRKDG 1231
QY	1407 KALTGEQTINGOHLFYKEDGHQVKGQVLTGDKVRYVDANSDDQAFNKSVTVNGKTYTF 1466
DB	1232 QLLTGEQTIIDGKVFQDNGVQVKGQGTATDASGLVRFVDRDQGHQ----- 1276
QY	1467 GNDGTAQTAGNPKGQTFKDGSDIRFYSMEGQVLTGSGYENAAQGMWLYV-KNGKVLTLGLQ 1525
DB	1277 -----VKGWYSTDSDNNWYVNESQVLTGLQ 1303
QY	1526 TVGSRVYVDENGIOAKGKAVRTSDGKIRYFPDENGSGSMITNWKVFGYGYFFGNDG 1582
DB	1304 TIDGQTVYFDDKGIQAKGKAVDENGSLRYFDADSGNMLRDKRWKNDGWNWYFNRNG 1360
RESULT 12	
JC5473	
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides	
C:Species: Leuconostoc mesenteroides	
C:Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004	
R:Accession: JC5473	
R:Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.	
Gene 182, 23-32, 1996	
A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leucon	
A:Reference number: JC5473; MUID:97136686; PMID:8982063	
A:Accession: JC5473	
A:Status: nucleic acid sequence not shown; translation not shown	
A:Molecule type: DNA	
A:Residues: 1-1290 <MON>	
A:Cross-references: UNIPROT:Q48756; UNIPARC:UPI000017ABDF; GB:U38181	
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose ont	
C:Genetics:	
C:Gene: darA	
C:Keywords: glycosyltransferase; hexosyltransferase	
F:78-870/Domain: catalytic #status predicted <CAT>	
F:922-1290/Domain: glucan-binding #status predicted <GB>	
Query Match 36.2%; Score 3021; DB 2; Length 1290;	
Best Local Similarity 44.7%; Pred. No. 2.4e-134;	
Matches 641; Conservative 210; Mismatches 412; Indels 172; Gaps 27;	
QY	151 MSNVKQVDGKYYYDQDGNVKNFAVSGDKIYYFD-ETGAYKDTSKVDADKSSSAVSQN 209
DB	1 MPNVKVDGKYVYFGDDGQPKNFITIIDGKPYFFDKTGALSNNDK-----QYVSEL 53

Db	1277	GSVKTGLVTLGKTYFSDYSVSQFNI-----NNNTYFQNDGTMQTGWNYGYVRYL	1331
Qy	1244	EANGQVKGDFTVSDGKLYF--YDVDSGDWMTDTFTEDKAGNWFYLGKGAAVTGAQTI	1301
Db	1332	NDSGIKWTG-WQTIDGNKYFDYFGAKTGIWNID-----GNYYFNNSGVMLTGWHOI	1383
Qy	1302	RGQKLYFKANGQQVKGDIVKGTGDKTIRYDAKSGEQVFNKTVKAAADGKTYVIGNDGVAVD	1361
Db	1384	NGSTYYFNSNGIANTGFIT--YLKCTYYFDSYGRMQIGSMTI---NGTSYFYFANGVM--	1436
Qy	1362	PSVVKGTGTFDASGALRPFYNLKGQLVTSQWYETAHNDWYVIO-SGKALTGEQTINGOHL	1420
Db	1437	-----KTSYDPSNTL-----AVGMVRDSYYQYVYLAAGTKLTGLQTDIGNTY	1479
Qy	1421	YFKEDGHQVKGQLVT-----GTDGKVRV-----YDANSNGDOAFNKSVTVNGKT	1463
Db	1480	YFDSNGIMQTG-IITINGNRYGFGVNGVMLYGLQFINNTYYNSYGISQTGFVTLGNT	1538
Qy	1464	YFQNDGTAQTA-----GNPKQGTPKDG--SDIRFYSMEG-----QLVTGSGWY	1505
Db	1539	YFDSYGEIRGLTYINNNTYYFNSKG-IMETGWSYLRVANPENGILLTGQTINGKTYV	1597
Qy	1506	ENAGQWLY-----VKNKVLTLGLOTVGSQRVYFDENGIOAKGKAVRTSDGKI	1553
Db	1598	FNSGSLYLDQYINGSYGFDKXGVMVLYGLQITGGNTYYLNSNGISQSG--FITLNGKT	1655
Qy	1554	RYFDENSG--SMITNQWKFVGYVYFGNDGARIYRGW	1589
Db	1656	YFDSYGYMETGIGN---INNYYFPGDNGT-LQTGW	1688

RESULT 14

A37184

glucan-binding protein - Streptococcus mutans

C:Species: Streptococcus mutans

C:Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 09-Jul-2004

C:Accession: A37184

R:Banas, J.A.; Russell, R.R.B.; Ferretti, J.J.

I:Infect. Immun. 58, 667-673, 1990

A:Title: Sequence analysis of the gene for the glucan-binding protein of Streptococcus mutans

A:Reference number: A37184; MUID:90170123; PMID:2307516

A:Accession: A37184

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-563 <BAN>

F:Cross-references: UNIPROT:Q54447; UNIPARC:UPI000008B8CB5; GB:M30945; NID:gl5366

F:169-188/Domain: cpl repeat homology <CP1>

F:264-283/Domain: cpl repeat homology <CP2>

F:349-368/Domain: cpl repeat homology <CP3>

F:504-523/Domain: cpl repeat homology <CP4>

F:525-548/Domain: cpl repeat homology <CP5>

Query Match		6.2%;	Score 518;	DB 2;	Length 563;
Best Local Similarity		30.6%;	Pred. No. 4.5e-17;		
Matches 158;		Conservative 59;	Mismatches 132;	Indels 168;	Gaps 21;

Qy	1118	ENGTAIRNTIYTDAGNSHYYANDGKRYENGYOQFGNDWRYFK-DGNMVAGLTVDGNVQ	1176
Db	146	EPATAAEN---NDAAPTNFGFKDGGK-----WYYKKADGQLATGMQIIDGKQL	190
Qy	1177	YFKDGVQAK-----DKI I-----VTRDGKRVFDQHNAGNAATNTFI----	1213
Db	191	YFNQDSQVKGELHVTGQIIIVHPIVSDSPSLVENVKIYFDPDSGELWKDRFYSSY	250
Qy	1214	AD-----KTGHWYVLGKDGVAVTGAQTVGKQKLYFEANGQOVKGDFTVSD-----	1258
Db	251	ADPLHYENIKHEGFFYLGEDGKAAIGWRITGGKKYIFDINGVQVKGLISTDGNVNLISQ	310
Qy	1259	EKGLYFYVDVSDGMMTDTFIEDK-----AG-----NMPYLGKGAAVTGAQIRGKL	1306
Db	311	KYGKSLFDDPTGEATNRFVNAKYFYFNPAGYVSTTDWFFYMGADGIVTDWQKIDGMDY	370

Search completed: February 11, 2006, 19:42:15
Job time : 44.8183 secs

Db 1712 PEIIVLNPNTF-HKKVNIINDSSSPYKWSSTEGSDFILVRYLEESNKKILOKIRI-KGIL 1769
QY 689 KQSDKGDA--RTSGVGVVMGNQPNFSLDGKVALNMGAAHNOEYRALMVSTKD 742
Db 1770 SNTQSFNKMISIDFKIKLGLSGYIMSNFKSFENSELDRDLGFK-----IIDNK- 1819
QY 743 GYATVATDADASKA-CLVXRTDBENGYLIFLNDLKGAVNPQVSGFLQVWVPV-GAADDQD 800
Db 1820 ---TYYYDESKLVKGLI--NINNSLFY-----DPIEFNLVTGWQTINGKKYYFD 1865
QY 801 IRVAASDTA--STDGKSLH--ODAAAMDSRVMEFSGPSFATKEEYTNVVIANNVDF 856
Db 1866 INTGAALTSYKLIINGKHYPFNNDGVNQLGV-FKPGDGFYFAPANTQNNI----- 1915
QY 857 VSWGITDFEMAPQYVSSSTDQFLDSYIQNGYAFTRDYDLGMSKANKYGTADQLVKAIK-- 914
Db 1916 -----EGQAI--VYQSKFLTLNGKKYYFDNNSKAVTGWRIINNEKY 1955
QY 915 -----ALHAKGLKVMADWPDQWYTPKQEVTVTRTDKFGKPIAGSQINHSLYVTDTKS 969
Db 1956 FNPNAIAAVGLQVIDN---NKYFNPDTAIIIS-----KGWQTVNGSRY---YFDTDTAI 2004
QY 970 SGDDYQAKYGGAFDLDELKYPPELFTKQISTGQAIIDPSVKIQWSA-----KYE-----N 1020
Db 2005 APNGYKTIDGKHYFD-----SDCVKLGVSFSTNGFEYFAPANTY 2045
QY 1021 GSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLGKVVESGIRYDGKGYIYN-----S 1076
Db 2046 NNIEGOAIVY-----QSKFLTLNGKKYYFDNN--SKAVTGWQIDSKYYFNTTABA 2097
QY 1077 ATGDQVKASFTIAGNLYFGKD-GWMTGAOTINGANYFFLENGTALNTYITDAQNS 1135
Db 2098 ATGWQ-----TIDGKKYYFNTTAAATGWQIDGKKYYFNTN-TAIASTGYTIINGKH 2150
QY 1136 HYIANDG-----KRYENGYO-----QFGND--- 1155
Db 2151 FYFNTDGMQIGVFGPGNGFEYFAPANTDANNIEGOAILYQNEFLTNGKKYYFGSDSKA 2210
QY 1156 ---MR-----YFKDGN--MAVGLTTVDGNVQYFDKQVQAKIIVTRDGKRYFD-- 1201
Db 2211 VTGWRIINNKYYFNPNAIAIHLCTINNDKYFYSYDGLQNGYITIERNN--FYFDAN 2268
QY 1202 -----QINGN-----AATNTFIADKTGHWYLLGKGVAVTG 1232
Db 2269 NESKMTVGVFGPGNGFEYFAPANTENNIEGOAIVYQNKFLT-LNGKKYYFDNDSKAVTG 2327
QY 1233 AQTGCKQKLYPEANGQVKGDFVTSDEGKLYFYDVSQDM-----TDTFI 1278
Db 2328 WQTDGKKYYFNLNTAEATGWQITID-GKKYYFNLNTAEATGWQITIDGKKYYFNTNTFI 2386
QY 1279 EDKAGNWFYLGKGAATGAOTIRGKLYFKANGQVKGDIIVKGTGDKIRY-----DAK 1333
Db 2387 -----ASTGYTSINGKHYPFNTDGMQIG-VFKPGNG-FEYFAPANTDAN 2429
QY 1334 SGEQ---VFNKTVKAADGKTYVIGNDGVAVDP-SVVGQOTFKDASGALRFYNLKGQLVG 1389
Db 2430 NIEGOAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGK-----KYFNTNTAVAV 2480
QY 1390 SGWYETANHDWYIOSGKAL--TGEQTINGQHLKFKEDGHQVKGQLVGTGDKVRY--- 1444
Db 2481 TGM-QTINGKKYYFNTNTSIASTGYTIIISGKHFYFNTDGMQIG-VFKGPDG-FEYFAPA 2537
QY 1445 --DANFGD-QAF---NKSVTVNGKTYFYFGNDGTAOTAGNPKQOTPKDGSIDRFYSMEGOL 1498
Db 2538 NTDANNIEGOAIRQNRFLYLHDNIYYFGNNSKAAT-----GWVTIDGN--RYF-FEPT 2589
QY 1499 VTSGWYENAOQWLYVXKQVLYTGLQTVGSRVYF-----DENGIOAKGKAVRTSD--- 1550
Db 2590 AMGANGYKTIIDKNKPYFRNGLPQIGVFKGNGFEYFAPANTDANNIE--GOAIRYQNRFL 2647
QY 1551 ---GKIRYFDENSGSMITNQKVFYGYQYFYFGNDGARIYRG 1588
Db 2648 HLLGKIYYFGNNSKAV--TGMQTINGKYYFYFMPDTAMAAAG 2686

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - proteih search, using sw model
Run on: February 11, 2006, 18:57:35 ; Search time 214.546 Seconds
(without alignments)
5228.676 Million cell updates/sec
Title: US-10-797-821-37
Perfect score: 8349
Sequence: 1 MEKNVRFKMHVKKRWTLTSL.....VYQYYFNGDGARIYRGMN 1590

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : UniProt_05.80.*
1: uniprot_spot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8349	100.0	1590	2	Q59983_9STRE	Q59983 streptococc
2	8256	98.9	1592	1	GTF2_STRDO	P27470 streptococc
3	8221	98.5	1590	2	Q55263_9STRE	Q55263 streptococc
4	7713.5	92.4	1597	1	GTF1_STRDO	P11001 streptococc
5	4663.5	55.9	1476	1	GTFB_STRMU	P09887 streptococc
6	4392.5	52.6	1455	1	GTFC_STRMU	P13470 streptococc
7	3636.5	43.6	1462	1	GTFD_STRMU	P49331 streptococc
8	3604	43.2	1577	2	Q54178_STRGN	Q54178 streptococc
9	3527	42.2	1575	2	Q9LCH3_STROR	Q9LCH3 streptococc
10	3502.5	42.0	1554	2	Q8KZL5_9STRE	Q8KZL5 streptococc
11	3485.5	41.7	1454	2	Q6PA94_LEUME	Q6PA94 leuconostoc
12	3464.5	41.5	1512	2	Q9WXJ5_9STRE	Q9WXJ5 streptococc
13	3410.5	40.8	1506	2	Q56CX8_9STRE	Q56CX8 streptococc
14	3325.5	39.8	1518	2	Q00600_STRSL	Q00600 streptococc
15	3282.5	39.3	2835	2	Q8G9Q2_LEUME	Q8G9Q2 leuconostoc
16	3245	38.9	1577	2	Q55265_STRSL	Q55265 streptococc
17	3214.5	38.5	1599	2	Q00599_STRSL	Q00599 streptococc
18	3128.5	37.5	1449	2	Q55264_STRSL	Q55264 streptococc
19	3114.5	37.3	1449	2	Q68542_STRSL	Q68542 streptococc
20	3088	37.0	1508	2	Q52224_LEUME	Q52224 leuconostoc
21	3087.5	37.0	1527	2	Q8KRE1_LEUME	Q8KRE1 leuconostoc
22	3084.5	36.9	1522	2	Q6TXVA_LEUME	Q6TXVA leuconostoc
23	3083	36.9	1508	2	Q9EZH5_LEUME	Q9EZH5 leuconostoc
24	3061.5	36.7	1527	2	Q9ZAR4_LEUME	Q9ZAR4 leuconostoc
25	3053	36.6	1477	2	Q9L466_LEUME	Q9L466 leuconostoc
26	3031	36.3	1365	1	GTF5_STRDO	P29136 streptococc
27	3025	36.2	1290	2	Q48756_LEUME	Q48756 leuconostoc
28	3003.5	36.0	1330	2	Q8ACN4_LEUME	Q8ACN4 leuconostoc
29	2994.5	35.9	1338	2	Q9WXJ4_9STRE	Q9WXJ4 streptococc
30	2974	35.6	1561	2	Q5SBM6_9LACO	Q5SBM6 lactobacill
31	2874	34.4	1463	2	Q5SBM6_LACFE	Q5SBM6 lactobacill

RESULT 1
Q59983_9STRE PRELIMINARY; PRT: 1590 AA.
AC Q59983;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfI;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OMZ176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
catalytic aspartic acid from two Streptococcus sobrinus alpha-
glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; 1HCX.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; F:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;
Query Match 100.0%; Score 8349; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKNVRFKMHVKKRWTLTSLVASATMLASALGASVASADTDTASDDSNQAVVTGQDTNN 60
Db 1 MEKNVRFKMHVKKRWTLTSLVASATMLASALGASVASADTDTASDDSNQAVVTGQDTNN 60
QY 61 QATDQTSIAATATSEQSASTDAATDQASAAEQTCQTASTDTAACTTTNANEAKVPTEN 120

Db 61 QATDQTSIAATATSEOSASTDAATDOASAAEQGTGTASTDAAQTNTTANAEKWPTE 120
QY 121 ENQGFTEMLAKNVATAESDSIPSDIAKMSNVKVDGKYYYYDODGNVKNFVAVSGD 180
Db 121 ENQGFTEMLAKNVATAESDSIPSDIAKMSNVKVDGKYYYYDODGNVKNFVAVSGD 180
QY 181 KIYYFDETCAYKDTSKVDADKSSAVSQNATIFAANNRAYSTSANKFEAVDNYLTADSWY 240
Db 181 KIYYFDETCAYKDTSKVDADKSSAVSQNATIFAANNRAYSTSANKFEAVDNYLTADSWY 240
QY 241 RPKSLDKGKTWESGKDDFRLLMAWPDTEKZNYNMKVGVIGIDKTYTAETSQADL 300
Db 241 RPKSLDKGKTWESGKDDFRLLMAWPDTEKZNYNMKVGVIGIDKTYTAETSQADL 300
QY 301 TAAABELVQARI BOKITSENNTKWLREAI SAFVKTQPMNGESEKPYDHLQNGALLFNQ 360
Db 301 TAAABELVQARI BOKITSENNTKWLREAI SAFVKTQPMNGESEKPYDHLQNGALLFNQ 360
QY 361 TDLTPTQSNVYLLNRTPTNQTGSLDSRFTYNPNDPLGGYDPLANDVNSPVVQAEOL 420
Db 361 TDLTPTQSNVYLLNRTPTNQTGSLDSRFTYNPNDPLGGYDPLANDVNSPVVQAEOL 420
QY 421 NWLHYLLNFGSIYANDADANFDSIRVDAVDNVDADLLQISSDYLKAAVIGIDKNNKANNH 480
Db 421 NWLHYLLNFGSIYANDADANFDSIRVDAVDNVDADLLQISSDYLKAAVIGIDKNNKANNH 480
QY 481 VSIWEASNDPTPYLHDDGDNLMNNDKPRLSMLSLAKPLDKRSGLNPLIHNSLVUREV 540
Db 481 VSIWEASNDPTPYLHDDGDNLMNNDKPRLSMLSLAKPLDKRSGLNPLIHNSLVUREV 540
QY 541 DDRETVTPSYSFARAHSEVODIIRDIKAEINPNSFGYFTOEIEQAQKYNEDLKK 600
Db 541 DDRETVTPSYSFARAHSEVODIIRDIKAEINPNSFGYFTOEIEQAQKYNEDLKK 600
QY 601 TDCKKTHNVPLSYLLLTNKGSIIPRVYVGMFTDDGGYMAKNTVNYDAIESLLKARMKY 660
Db 601 TDCKKTHNVPLSYLLLTNKGSIIPRVYVGMFTDDGGYMAKNTVNYDAIESLLKARMKY 660
QY 661 VSGQAMQNYQINGEILLTSVRYGKALKQSDKGDATTGSGVYVGMQNPNSLDGKV 720
Db 661 VSGQAMQNYQINGEILLTSVRYGKALKQSDKGDATTGSGVYVGMQNPNSLDGKV 720
QY 721 ALNMGAAHANQBYRALMYSTKDGVAITYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
Db 721 ALNMGAAHANQBYRALMYSTKDGVAITYATDADASKAGLVKRTDENGILYFLNDDLKGVAN 780
QY 781 PQVSGFLOVWPVGAADDQDQIRVAASDTASTDGKSLHODAAVMSRVMEFSGNFOSFATK 840
Db 781 PQVSGFLOVWPVGAADDQDQIRVAASDTASTDGKSLHODAAVMSRVMEFSGNFOSFATK 840
QY 841 EEEYTNVVIANNVDFKVSNGITDFEMAPQYVSSTDPGQLDSVIQNGYAFTRDYDLGMSKA 900
Db 841 EEEYTNVVIANNVDFKVSNGITDFEMAPQYVSSTDPGQLDSVIQNGYAFTRDYDLGMSKA 900
QY 901 NKYGTAQDLVKAIALHAKGLKVMADWPDQMYTFPKQEVVTVTRTDKFKPIAGSQINH 960
Db 901 NKYGTAQDLVKAIALHAKGLKVMADWPDQMYTFPKQEVVTVTRTDKFKPIAGSQINH 960
QY 961 SLVYVTDTKSSGDDYQAKYGGAFDELKEKYPELFTKKQISTGQALDPSVKIKQMSAKYFN 1020
Db 961 SLVYVTDTKSSGDDYQAKYGGAFDELKEKYPELFTKKQISTGQALDPSVKIKQMSAKYFN 1020
QY 1021 GSNILGRGADYVLSQVSNKYNFVASDTLFLPSSLLGKVKVSGIRYDGKGIYNSSATGD 1080
Db 1021 GSNILGRGADYVLSQVSNKYNFVASDTLFLPSSLLGKVKVSGIRYDGKGIYNSSATGD 1080
QY 1081 QVKASFITEAGNLYYFGKDGVMVTCQAOTINGANYEFLENGTALRNTIYTDAGNSHYAN 1140
Db 1081 QVKASFITEAGNLYYFGKDGVMVTCQAOTINGANYEFLENGTALRNTIYTDAGNSHYAN 1140
QY 1141 DGKRYENGYQQFGNDWRVFKDGNMAVGLTTVDGNVQYFDKQVQAKDKIIVTRDCKVRYF 1200
Db 1141 DGKRYENGYQQFGNDWRVFKDGNMAVGLTTVDGNVQYFDKQVQAKDKIIVTRDCKVRYF 1200

QY 1201 DOHNGNAATNTFIADKTGHWTYLLGKDGVAVTGAQTVGKQKLYFEANGQOVKGDFTVSDBG 1260
Db 1201 DOHNGNAATNTFIADKTGHWTYLLGKDGVAVTGAQTVGKQKLYFEANGQOVKGDFTVSDBG 1260
QY 1261 KLYFYDVDSGDMWTDTFIEDKAGNWFYLGKDGAAVTGAQTIRGOKLYFKANGQOVKGDIV 1320
Db 1261 KLYFYDVDSGDMWTDTFIEDKAGNWFYLGKDGAAVTGAQTIRGOKLYFKANGQOVKGDIV 1320
QY 1321 KGTGDKIRYIDAKSGEQVFNKTVKAAADKTYVIGNDGVAVDPSVVYKGTQFKDASGALRFY 1380
Db 1321 KGTGDKIRYIDAKSGEQVFNKTVKAAADKTYVIGNDGVAVDPSVVYKGTQFKDASGALRFY 1380
QY 1381 NLKQOLVTGSGWYETANHDWYIIOGSKALTGEBOTINGOHLYPKEDGHQVKGOLVTGTGDK 1440
Db 1381 NLKQOLVTGSGWYETANHDWYIIOGSKALTGEBOTINGOHLYPKEDGHQVKGOLVTGTGDK 1440
QY 1441 VRYDANSDDQAFNKSVTVNGKTYVYFGNDGTAQTAGNPKGOTFKDGSDIRFYSMEGQLVT 1500
Db 1441 VRYDANSDDQAFNKSVTVNGKTYVYFGNDGTAQTAGNPKGOTFKDGSDIRFYSMEGQLVT 1500
QY 1501 GSGWYENAAQGWLYVYKNGKVLTLGTLQTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDENS 1560
Db 1501 GSGWYENAAQGWLYVYKNGKVLTLGTLQTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDENS 1560
QY 1561 GSMITNQNWKVYQYVYFGNDGARIYRGWN 1590
Db 1561 GSMITNQNWKVYQYVYFGNDGARIYRGWN 1590

RESULT 2

GTFF2_STRDO STANDARD; PRT; 1592 AA.
ID GTFF2_STRDO
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX STRAIN=6715 / Serotype G;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 16 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----

DR	EMBL; D90213; BAA14241.1; -; Genomic_DNA.	
DR	HSP; P06653; 1GVM.	
DR	InterPro; IPR002479; Cell wall bd put.	
DR	InterPro; IPR003318; Glyco_hydro_70.	
DR	Pfam; PF01473; CW_binding_1; 3.	
DR	Pfam; PF02324; Glyco_hydro_70; 1.	
KW	Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.	
FT	SIGNAL 1 38 Potential.	
FT	CHAIN 39 1592	
FT	REPEAT 1093 1142 1.	
FT	REPEAT 1158 1207 2.	
FT	REPEAT 1222 1272 3.	
FT	REPEAT 1287 1337 4.	
FT	REPEAT 1402 1451 5.	
FT	REPEAT 1514 1563 6.	
FT	REPEAT 1577 1592 7 (incomplete).	
FT	REGION 39 1044 Catalytic (approximate).	
FT	REGION 1093 1592 7 X tandem repeats.	
FT	REGION 1093 1592 Glucan-binding (approximate).	
SQ	SEQUENCE 1592 AA; 176168 MW; BC0A66D07931ECF CRC64;	
Query Match 98.9%; Score 8256; DB 1; Length 1592;		
Best Local Similarity 98.9%; Pred. No. 0;		
Matches 1574; Conservative 6; Mismatches 10; Indels 2; Gaps 1;		
QY	1 MEKNVRFPQHKVKKRWVTLVSASATMLASALGASVASADTDTASDDSNQAVVTGDDTTNN 60	
DB	1 MEKNVRFPQHKVKKRWVTLVSASATMLASALGASVASADTDTASDDSNQAVVTGDDTTNN 60	
QY	61 QATDQTSIAATATSEOSASTDAATDOASAAEQTGTASTDTAAQTTTNANEAKVVPTE 120	
DB	61 QATDQTSIAATATSEOSASTDAATDOASAAEQTGTASTDTAAQTTTNANEAKVVPTE 120	
QY	121 ENQGFDEMLAKNAVATAESIPSDLAKSNVQVGGKYYYDQDGNVKNFAVSVD 180	
DB	121 ENQGFDEMLAKNAVATAESIPSDLAKSNVQVGGKYYYDQDGNVKNFAVSVD 180	
QY	181 KIYFDETCAYKDTSKVDADKSSAVSQNATIFAANNRAYSTSAKNFEADVNTADSWY 240	
DB	181 KIYFDETCAYKDTSKVDADKSSAVSQNATIFAANNRAYSTSAKNFEADVNTADSWY 240	
QY	241 RPKSILKDGKWTESGKDFRPLLMAMPDPTETKRYVNMKNVGVGIDKITYTAETSQADL 300	
DB	241 RPKSILKDGKWTESGKDFRPLLMAMPDPTETKRYVNMKNVGVGIDKITYTAETSQADL 300	
QY	301 TAAAEVLQARIQKITSENNTKWLREAI SAFVKTPQWNGESEKPYDDHLQNGALLFDNQ 360	
DB	301 TAAAEVLQARIQKITSENNTKWLREAI SAFVKTPQWNGESEKPYDDHLQNGALLFDNQ 360	
QY	361 TDLTPTQSNRYLLNRTPTNQTGSLDSRFTYNPNDPLGGYDFLLANDVDNSNPVVAEOL 420	
DB	361 TDLTPTQSNRYLLNRTPTNQTGSLDSRFTYNPNDPLGGYDFLLANDVDNSNPVVAEOL 420	
QY	421 NMLHYLLNFGSIYANDADANFDSIRVDAVDNVDADQLQISSDYLKAAAYGIDKNNKNNNH 480	
DB	421 NMLHYLLNFGSIYANDADANFDSIRVDAVDNVDADQLQISSDYLKAAAYGIDKNNKNNNH 480	
QY	481 VSIVEAWSNDPTPYLHDDGDNLMMNDKPERLSMLSLAKPLDKRSGLNPLIHSNLDREV 540	
DB	481 VSIVEAWSNDPTPYLHDDGDNLMMNDKPERLSMLSLAKPLDKRSGLNPLIHSNLDREV 540	
QY	541 DDREVEVTPSYSFARAHSEVDODIIRDIKASINPNSFGYSTQEEIEQAFKIYNEDLKK 600	
DB	541 DDREVEVTPSYSFARAHSEVDODIIRDIKASINPNSFGYSTQEEIEQAFKIYNEDLKK 600	
QY	601 TDKKYTHNVPLSYTLTLTNKGSIPRVYVYGMFTDDGGYMANKTNYDAIESLLKARMKY 660	
DB	601 TDKKYTHNVPLSYTLTLTNKGSIPRVYVYGMFTDDGGYMANKTNYDAIESLLKARMKY 660	
QY	661 VSGGQAMQNYQINGEILTSVRYGKALKQSDGKDATTRTSVGVMGNQPNFSLDGKV 720	
DB	661 VAGGQAMQNYQINGEILTSVRYGKALKQSDGKDATTRTSVGVMGNQPNFSLDGKV 720	
QY	721 ALNMGAAHANQBYRALMWSTKDGVAITYATDADASKAGLVKRTDENGLYFLNDDLKGVAN 780	
DB	721 ALNMGAAHANQBYRALMWSTKDGVAITYATDADASKAGLVKRTDENGLYFLNDDLKGVAN 780	
QY	781 PQVSGFLQWVVPVGAADDQDQIRVAASDTRASTDGSLSLHQAAMDSRVWFGSFPQSFATK 840	
DB	781 PQVSGFLQWVVPVGAADDQDQIRVAASDTRASTDGSLSLHQAAMDSRVWFGSFPQSFATK 840	
QY	841 EBEYTNVVIANNVDKFSWGIITDFEWAPOVYSTDQGLDSTVIQNGYATFTDRYDLGWSKA 900	
DB	841 EBEYTNVVIANNVDKFSWGIITDFEWAPOVYSTDQGLDSTVIQNGYATFTDRYDLGWSKA 900	
QY	901 NKYGTADQLVKAIKALHAKGLKVMADWPDQMYTTPKQEVVTVTRTDKFKPIAGSQINH 960	
DB	901 NKYGTADQLVKAIKALHAKGLKVMADWPDQMYTTPKQEVVTVTRTDKFKPIAGSQINH 960	
QY	961 SLIYVDTKSSGDDYQAKYGAGFLDELKEYPELFTKKQISTGOAIDPSPVKIKQWSAKYFN 1020	
DB	961 SLIYVDTKSSGDDYQAKYGAGFLDELKEYPELFTKKQISTGOAIDPSPVKIKQWSAKYFN 1020	
QY	1021 GSNILGRGADYVLSQVSNKYFNVAASDTLFLPSSLLGKVVESGIRYDGGKGIYNSATGD 1080	
DB	1021 GSNILGRGADYVLSQVSNKYFNVAASDTLFLPSSLLGKVVESGIRYDGGKGIYNSATGD 1080	
QY	1081 QVKASFIIEAGNLYYFGKDGVMVTGAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140	
DB	1081 QVKASFIIEAGNLYYFGKDGVMVTGAQTINGANYFFLENGTALRNTIYTDAGNSHYAN 1140	
QY	1141 DGKRY- -ENGVOQFGNDWRYFKDGNMAGLTTVDGNVQYFDKDGVOAKDKIIVTRDGKVR 1198	
DB	1141 DGKRYENENGVOQFGNDWRYFKDGNMAGLTTVDGNVQYFDKDGVOAKDKIIVTRDGKVR 1200	
QY	1199 YFDOHNGNAATNTEFADTKGHWYILGKGVAVTGAQTVGKQKLYFEANGQVKGDFVTS 1258	
DB	1201 YFQHNAGVNTNTEFADTKGHWYILGKGVAVTGAQTVGKQKLYFEANGQVKGDFVTS 1260	
QY	1259 EGKLYFYDSDGMDMTDTFIEDKAGNWFYLGKDGAAVTAQTIHQKLYFKANGQOVKGD 1318	
DB	1261 EGKLYFYDSDGMDMTDTFIEDKAGNWFYLGKDGAAVTAQTIHQKLYFKANGQOVKGD 1320	
QY	1319 IVKGTGDKIRYDAKSGOVFNKTKAADGKTYVIGNDGVAVDPSVVGQTFKDSAGALR 1378	
DB	1321 IVKGTGDKIRYDAKSGOVFNKTKAADGKTYVIGNDGVAVDPSVVGQTFKDSAGALR 1380	
QY	1379 FYNLAGQLVTGSGWYETANHDWVYIQSGKALTGEQTINGOHLVFKEDGHOVKQLVTGTD 1438	
DB	1381 FYNLAGQLVTGSGWYETANHDWVYIQSGKALTGEQTINGOHLVFKEDGHOVKQLVTGTD 1440	
QY	1439 GKRVYYDANSQDAFNKSVTVNGKTYFFGNDGTAQTAGNPKQTFKQSDIRFYSMEGOL 1498	
DB	1441 GKRVYYDANSQDAFNKSVTVNGKTYFFGNDGTAQTAGNPKQTFKQSDIRFYSMEGOL 1500	
QY	1499 VTGSGWYENAGQWLYVKNKVLTLGLQTVGSGORVYFDENGIQAKGKAVRTSDGKIRYFDE 1558	
DB	1501 VTGSGWYENAGQWLYVKNKVLTLGLQTVGSGORVYFDENGIQAKGKAVRTSDGKIRYFDE 1560	
QY	1559 NSGSMITNQWKEVYQYVYFGNDGARIYRGN 1590	
DB	1561 NSGSMITNQWKEVYRYYFNGDGIYRGN 1592	
RESULT 3		
Q55263 9STRE		
ID	Q55263_9STRE PRELIMINARY; PRT; 1590 AA.	
AC	Q55263;	
DT	01-NOV-1996 (TRENBLrel. 01, Created)	
DT	01-JAN-1998 (TRENBLrel. 05, Last sequence update)	
DT	01-JUN-2003 (TRENBLrel. 24, Last annotation update)	
DE	GTF-I.	
GN	Name=Glucosyltransferase;	
OS	Streptococcus sobrinus.	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OC	Streptococcus.	

OX	NCBI_TaxId=1310;	
RN	[1] NUCLEOTIDE SEQUENCE.	
RP	STRAIN=ATCC 33478;	
RC	Sato S.;	
RA	"DNA and amino-acid sequences of water-insoluble-glucon synthetase	
RT	produced from <i>Streptococcus sobrinus</i> ATCC 33478.";	
RL	Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).	
RN	[2] PROTEIN SEQUENCE.	
RP	MEDLINE=91224988; PubMed=1827439;	
RX	Mooser G., Hetta S.A., Paxton R.J., Shively J.E., Lee T.D.;	
RA	"Isolation and sequence of an active-site peptide containing a	
RT	catalytic aspartic acid from two <i>Streptococcus sobrinus</i> alpha-	
RT	glucosyltransferases.";	
RL	J. Biol. Chem. 266:8916-8922(1991).	
DR	EMBL; D63570; BAA09792.1; -; Genomic_DNA.	
DR	PIR; A39841; A39841.	
DR	HSSP; P06653; 1GVN.	
DR	GO; GO:0009250; P-glucon biosynthesis; IEA.	
DR	InterPro; IPR002479; CW_binding.	
DR	InterPro; IPR003318; Glyco_hydro_70.	
DR	Pfam; PF01473; CW_binding_1; 3.-	
DR	Pfam; PF02324; Glyco_hydro_70; 1.	
SQ	SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;	
Query Match 98.5%; Score 8221; DB 2; Length 1590;		
Best Local Similarity 98.5%; Pred. No. 0;		
Matches 1566; Conservative 9; Mismatches 15; Indels 0; Gaps 0;		
QY	1 MEKNVRFKWHKVKRWVTLVSASATMLASALGASVASADTDPASDSSNOAVVTGQTTNN	60
DB	1 MEKNVRFKWHKVKRWVTLVSASATMLASALGASVASADTDPASDSSNOAVVTGQTTNN	60
QY	61 QATDQTSIAATATSEQSASTDAATDQASAAEQTTASTDTAAQTNNANEAKVPTEN	120
DB	61 QATDQTSIAATATSEQSASTDAATDQASAAEQTTASTDTAAQTNNANEAKVPTEN	120
QY	121 ENQGTDEMLAKAVATAESDSIPSDLAKMSNVQVQKYYVDQDGNVKNFAVSVD	180
DB	121 ENQGTDEMLAKAVATAESDSPSPDLAKMSNVQVQKYYVDQDGNVKNFAVSVD	180
QY	181 KIYVDETGAYKDTSKVDADKSSAVSQNATIFAAANNRAYSTSAKNFEAVDNYLTADSWY	240
DB	181 KIYVDETGAYKDTSKVDADKSSAVSQNATIFAAANNRAYSTSAKNFEAVDNYLTADSWY	240
QY	241 RPKSILKDGKTWTEGKDDFRPLLMAWMPDTEKKNYNNMKVYVGIKTYTAETSQADL	300
DB	241 RPKSILKDGKTWTEGKDDFRPLLMAWMPDTEKKNYNNMKVYVGIKTYTAETSQADL	300
QY	301 TAAAEVLQARIRBKITSENNTKWLREAIISAFVKTPQWNGESEKPYDDHLQNGALLFDNQ	360
DB	301 TAAAEVLQARIRBKITSENNTKWLREAIISAFVKTPQWNGESEKPYDDHLQNGALLFDNQ	360
QY	361 TDLTPDQTSNRYLLNRTPTNQTGSLDSRFTYNNPNDPLGQYDFLLANDVNSNPVQASOL	420
DB	361 TDLTPDQTSNRYLLNRTPTNQTGSLDSRFTYNNPNDPLGQYDFLLANDVNSNPVQASOL	420
QY	421 NWLHYLLNFGSIYAKDADANFDSIRVDVNDVADLLQISSDYLKAAAGIDKNNKNANH	480
DB	421 NWLHYLLNFGSIYAKDADANFDSIRVDVNDVADLLQISSDYLKAAAGIDKNNKNANH	480
QY	481 VSIWEASGNDPTPYLHDDGDNLMNDKFRLSMLWSLAKPLDKRSGLNPLIHNSLVDRREV	540
DB	481 VSIWEASGNDPTPYLHDDGDNLMNDKFRLSMLWSLAKPLDKRSGLNPLIHNSLVDRREV	540
QY	541 DDREVTVPVSFAFAHSEVDIIRDIIRKABINPNSFGYPTQEEIEQAFKIYNEDLKK	600
DB	541 DDREVTVPVSFAFAHSEVDIIRDIIRKABINPNSFGYPTQEEIEQAFKIYNEDLKK	600
QY	601 TDKKYTHYNVPLSYTLTLTNKGSIPRVYVYGMFTDDGQYMANKTNYDAIESLLKARMKY	660
DB	601 TDKKYTHYNVPLSYTLTLTNKGSIPRVYVYGMFTDDGQYMANKTNYDAIESLLKARMKY	660

RESULT 4

ID	GTFL_STRDO	STANDARD;	PRT;	1597 AA.
AC	P11001;			
DT	01-JUL-1989	(Rel. 11, Created)		
DT	01-JUL-1989	(Rel. 11, Last sequence update)		
DT	13-SEP-2005	(Rel. 48, Last annotation update)		

QY	661 VSGGQAMQNYOINGEILLTSVRYGKALKQSDKGDATTTRTSGVGVVMGNQNFSLDGKVV	720
DB	661 VSGGQAMQNYOINGEILLTSVRYGKALKQSDKGDATTTRTSGVGVVMGNQNFSLDGKVV	720
QY	721 ALNMGAAHANQYRALMVSTKDGVAATATDADASKAGLTKRTDENGYYLFLNDDLKGVAN	780
DB	721 ALNMGAAHANQYRALMVSTKDGVAATATDADASKAGLTKRTDENGYYLFLNDDLKGVAN	780
QY	781 PQVSGFLQVVPVGAADDQD1RVAASDTASTDGSLSLHQAADMSRVMEFEGSNFOSFATK	840
DB	781 PQVSGFLQVVPVGAADDQD1RVAASDTASTDGSLSLHQAADMSRVMEFEGSNFOSFATK	840
QY	841 EEEYTNVVIANNVOKFYSGWITDEMAPOQYVSSYSDGQFLDSVIQNGYAFTRDYDLGMSKA	900
DB	841 EEEYTNVVIANNVOKFYSGWITDEMAPOQYVSSYSDGQFLDSVIQNGYAFTRDYDLGMSKA	900
QY	901 NKYGTADOLVKAIKALHAKGLKVMADVPDQMYTFPKQEVVTVTRTDKFGKPIAGSQINH	960
DB	901 NKYGTADOLVKAIKALHAKGLKVMADVPDQMYTFPKQEVVTVTRTDKFGKPIAGSQINH	960
QY	961 SLYVTDTKSGDDYQAKYGGAFDLDELKEXYDELFTKKQISTGQAIIDPSVKIKQWSAKYFN	1020
DB	961 SLYVTDTKSGDDYQAKYGGAFDLDELKEXYDELFTKKQISTGQAIIDPSVKIKQWSAKYFN	1020
QY	1021 GSNILGRGADVVLSDOVSNNKYFNVASDTLFLPSSILGKVVESGIRYDCKGYIYNSSATGD	1080
DB	1021 GSNILGRGADVVLSDOVSNNKYFNVASDTLFLPSSILGKVVESGIRYDCKGYIYNSSATGD	1080
QY	1081 QVKASFITEAGNLVYFGKDGVMVTGAOTINGANTFFLENGTALRNTIYTDAGQNSHYVAN	1140
DB	1081 QVKASFITEAGNLVYFGKDGVMVTGAOTINGANTFFLENGTALRNTIYTDAGQNSHYVAN	1140
QY	1141 DGKRYENGYQQFGNDWRVYFKDGNMAVGLTTVDGNVQYFDKDGVAQDKIIIVTRDGKVRYP	1200
DB	1141 DGKRYENGYQQFGNDWRVYFKDGNMAVGLTTVDGNVQYFDKDGVAQDKIIIVTRDGKVRYP	1200
QY	1201 DOHNGNAATNTFIADKTGHVYLLGKDGVAVTAQTVGKQKLYFEANGQVKGDFVTSDEG	1260
DB	1201 DOHNGNAATNTFIADKTGHVYLLGKDGVAVTAQTVGKQKLYFEANGQVKGDFVTSDEG	1260
QY	1261 KLYFYDVSQDMWTDFTIEDKAGNFWLKGKGAATVGAQTIRGOKLYFKANGQQVKGDIV	1320
DB	1261 KLYFYDVSQDMWTDFTIEDKAGNFWLKGKGAATVGAQTIRGOKLYFKANGQQVKGDIV	1320
QY	1321 KGTGDKIRYDAKSGEQVFNKTVAADGKTYYIGNNGVAVDPSVVVKGQTFPKDASGALRPF	1380
DB	1321 KGTGDKIRYDAKSGEQVFNKTVAADGKTYYIGNNGVAVDPSVVVKGQTFPKDASGALRPF	1380
QY	1381 NLKGOLVTGSGWYETANHDWYIOSGKALTGEOTINGOHLYPKEDGHQVKGOLVTRTDGK	1440
DB	1381 NLKGOLVTGSGWYETANHDWYIOSGKALTGEOTINGOHLYPKEDGHQVKGOLVTRTDGK	1440
QY	1441 VRYIDANSQDAFNKSVTVNGKTYVFGNDGTAAQTAGNPKGQTFKDGSDIRFYSMEGQLVT	1500
DB	1441 VRYIDANSQDAFNKSVTVNGKTYVFGNDGTAAQTAGNPKGQTFKDGSDIRFYSMEGQLVT	1500
QY	1501 GSGWYENAAQGWLVYXGKVLTLGLQTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDENS	1560
DB	1501 GSGWYENAAQGWLVYXGKVLTLGLQTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDENS	1560
QY	1561 GMSMITNOMKPYQYQYFGNDGARIYRCWN	1590
DB	1561 GMSMITNOMKPYQYQYFGNDGARIYRCWN	1590


```

Db 1261 FWTADGKLYFDVDSGDMWNTFTEDKAGNWFYLGKGAATVGAQTIKGQKLYFKANGQ 1320
QY 1314 QVKGDIIVKGTGKIRYYDAKSGEQVFNKTVKAAADGKTVYVIGNDGVAVDPSVVVKGOTFKDA 1373
Db 1321 QVKGDIIVKADGKIRYYDAQTGEQVFNKSV-SVNGKTYVFGSDGTAQTQANPKGQTFKDG 1379
QY 1374 SCALFPYNLKGOLVTGSGWYETANDHWYIQSGKALTGEOTINGOHLYFKEDGHQVKGL 1433
Db 1380 SGVLFPYNLEGGYVSGSGWYETAEHEWVYVKSGLVLTGAQTIGNORVYFKONGHQVKGL 1439
QY 1434 VTGTGKRYRYDANGSDQAFNKSVTVNGKTYVFGNDGTAQTAGNPKGOTFKDGSDI-RFY 1492
Db 1440 VTGNDGKLYRYDANGSDQAFNKSVTVNGKTYVFGSDGTAQTQANPKGQTFKDGSGVLRPY 1499
QY 1493 SMEGOLVTGSGWYENAOQGLVYKNGKVLTLGLQTVGSRVYPDENGIOAKGKAVRTSDGK 1552
Db 1500 NLEGGYVSGSGWYKNAQGLVYKNGKVLTLGLQTVGNQKVVYFDKNGIOAKGKAVRTSDGK 1559
QY 1553 TRYFDENSGSMITNQKVFYGGYVFGNDGARIYRGWN 1590
Db 1560 VRYFDENSGSMITNQKVFYGGYVFGSDGAAYVRGWN 1597

```

RESULT 5

```

GTFB_STRMU
ID GTFB_STRMU STANDARD; PRT; 1476 AA.
AC P08987; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (BC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfB; OrderedLocusNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype C;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX STRAIN=UAI159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Prineaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.

```

```

CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-S1 synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 10 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M17361; AAA88588.1; -; Genomic_DNA.
CC EMBL; D88651; BAA26101.1; -; Genomic_DNA.
CC EMBL; D88654; BAA26105.1; -; Genomic_DNA.
CC EMBL; D88657; BAA26109.1; -; Genomic_DNA.
CC EMBL; D88660; BAA26113.1; -; Genomic_DNA.
CC EMBL; D89577; BAA26119.1; -; Genomic_DNA.
CC EMBL; AE014940; AAN58705.1; -; Genomic_DNA.
CC PIR; B33135; B33135.
CC HSSP; P06653; 1H8G.
CC InterPro; IPR002479; Cell wall bd put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_I; 4.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Complete proteome; Dental caries; Glucosyltransferase; Repeat; Signal;
CC Transference.
KW SIGNAL 1 34 Potential.
FT CHAIN 35 1476 Glucosyltransferase-I.
FT REPEAT 1097 1130 A repeat.
FT REPEAT 1161 1210 1.
FT REPEAT 1225 1275 2.
FT REPEAT 1290 1340 3.
FT REPEAT 1355 1405 4.
FT REPEAT 1420 1470 5.
FT REGION 35 1051 Catalytic (approximate).
FT REGION 1097 1476 Glucan-binding (approximate).
FT REGION 1161 1470 S X tandem repeats.
FT VARIANT 62 62 S -> T (in strain MT4239).
FT VARIANT 65 65 T -> I (in strain GS-5).
FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245,
FT strain MT4251, strain MT4467 and strain
FT MT8148).
FT VARIANT 78 78 Q -> P (in strain MT4251).
FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245,
FT strain MT4251, strain MT4467 and strain
FT MT8148).
FT VARIANT 89 89 S -> F (in strain MT4251).
FT VARIANT 168 168 K -> N (in strain MT4251).
FT VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and
FT strain MT8148).
FT VARIANT 399 399 N -> R (in strain MT4239).
FT VARIANT 474 474 I -> T (in strain MT4239).
FT VARIANT 512 512 K -> R (in strain MT8148).
FT VARIANT 519 519 F -> Y (in strain MT8148).
FT VARIANT 701 701 T -> I (in strain MT8148).
FT VARIANT 708 708 A -> V (in strain MT8148).
FT VARIANT 938 938 F -> L (in strain MT8148).
FT VARIANT 952 957 FGKPEV -> YGTPVA (in strain GS-5, strain
FT MT4239 and strain MT4467).
FT VARIANT 963 964 SV -> NT (in strain GS-5, strain MT4239
FT and strain MT4467).
FT VARIANT 968 970 ADS -> VDG (in strain GS-5, strain MT4239
FT and strain MT4467).
FT VARIANT 1086 1086 A -> T (in strain MT4239).
FT VARIANT 1158 1158 S -> N (in strain MT4239).
FT VARIANT 1163 1163 H -> Y (in strain MT4251).
FT VARIANT 1168 1168 E -> K (in strain MT8148).
FT VARIANT 1182 1182 Y -> C (in strain MT8148).
FT VARIANT 1234 1234 A -> P (in strain MT4239).
FT VARIANT 1263 1263 R -> H (in strain GS-5 and strain
FT MT4467).

```



```
FT VARIANT 1263 1263 R -> P (in strain MT8148).
FT VARIANT 1264 1264 Y -> H (in strain GS-5, strain MT4239,
FT VARIANT 1272 1272 S -> G (in strain MT8148).
FT VARIANT 1272 1272 S -> G (in strain GS-5, strain MT4239,
FT VARIANT 1329 1329 H -> Y (in strain MT8148).
FT VARIANT 1329 1329 H -> Y (in strain GS-5 and strain
FT VARIANT 1394 1394 MT4467).
FT VARIANT 1394 1394 Y -> H (in strain GS-5, strain MT4239,
FT VARIANT 1402 1402 S -> G (in strain MT8148).
FT VARIANT 1402 1402 S -> G (in strain GS-5, strain MT4239,
FT VARIANT 1459 1459 Y -> H (in strain MT8148).
FT VARIANT 1459 1459 Y -> H (in strain MT4467).
FT CONFLICT 570 570 R -> A (in Ref. 1).
FT CONFLICT 800 817 ADOQVRAASTAPSTDGK -> LIKMFALRLARPHQOMA
FT CONFLICT 1310 1310 (in Ref. 1).
FT CONFLICT 1476 1476 H -> L (in Ref. 1).
FT SEQUENCE 1476 AA; 165847 MW; 9C6E09F731B4CBFC CRC64;

Query Match 55.9%; Score 4663.5; DB 1; Length 1476;
Best Local Similarity 57.1%; Pred. No. 5.5e-207;
Matches 906; Conservative 206; Mismatches 339; Indels 135; Gaps 15;

QY 1 MEKVREKHKVKKRWVTLVSASATWLASALGASVASADTDTA-----SDSNQAVVTGQ 56
DB 1 MDKVKRYKKRWVTVSVASAVMTLTLTSLGLVKADSNESKSIQISNDSTSVVTANE 60
QY 57 TTNOQATDTSIAATATSBQSASTDAATQASAEQTQGTASTDTAAQTITNANEAKW 116
DB 61 ESN--VTTEVTSKQEAASQTNHTVTIISSTSVNPKVSNPYTVGETASNGEKL--- 115
QY 117 PTENENQGTDEMLAA-----KNVATAESQIPSD-----LAKMSNVKQVQDKYYY 163
DB 116 ----QNOTTVTKTSEAAANNISKQTEADTDVIDDSNAANLQILEKLPNVKEIDKYYY 171
QY 164 YDQDGNVKNKFAVSVDKIYDETCAYKDTSKVDADKSSAVSQNATIPAAANNRAYSTS 223
DB 172 YDNGKVRNTFLIADGKILHDETCAYTDTSIDTVNK--DIVTTRSNLYKKNQVYDRS 229
QY 224 AKNFVADNYLTADSWYRPSKILKQGTWTEGKODFRPLLMANWPDTEKRYNVYMNK 283
DB 230 AQSFVHVHLYTAESWYRPSKILKQGTWTEGKODFRPLLMANWPDTEKRYNVYMNK 289
QY 284 VVGIDKTYTAETSQADLTAAELVQARIEQKITSNNKWLREAI SAFVKTQPMNGESE 343
DB 290 QLGINKTYDTSNQLNATAATIAQIEAKITLKNLTDWLAQTISAFVKTQSAWNSDSE 349
QY 344 KPVDHLQNGALLFNQDNLTDLPDTSQNYLLNRTPNQTSGLDSRFTYNDPLGQYDPL 403
DB 350 KPFDHLQNGAVLYDNEGKLTTPYANSYRILNRTPTNQTGKKDPRYT--ADNTIGYEPL 407
QY 404 LANDVDNSPVVQAEQNLWHLNFGSIYANDADANFDSIRVDAVNDVADLLQIISDY 463
DB 408 LANDVDNSPVVQAEQNLWHLNFGSIYANDADANFDSIRVDAVNDVADLLQIISDY 467
QY 464 LKAAVYGIQKNNKANNHVSIVSAWSDNDTPYLHDDGDLNMDNKNFRLSMLSLAKPLDK 523
DB 468 LKAAAGIHKNDKAAANDHLSILEASDNDTPYLHDDGDLNMDNKNFRLSMLSLAKPLNQ 527
QY 524 RSLGNPLHNSLVDRVDREVEVTPSYSFARAHDSSEVODIIRDIKAEINPNPSFGYPT 583
DB 528 RSGMPLHNSLVDRVDREVEVTPSYSFARAHDSSEVODIIRDIKAEINPNPSFGYPT 587
QY 594 OBEIEQAFKIYNEDLKTKTKYTHYNPLSYLLLTNKGSIIPRVYVGMFTDDGOYMANK 643
DB 598 MEEIKAEIYNKDLATEKKYTHNTALSVALLTNKGSIIPRVYVGMFTDDGOYMAHK 647
QY 644 TVNYDAIESLLKARMKYVSGGQAMQYQIINGEILTSVRYGKALKQSKDGATRTTSGV 703
DB 648 TINYEAETLLKARIKYVSGGQAMRQVGNSEIITSVRYGKALKATDTGDRTRTSGV 707
QY 704 GYVMGNQNFSLDGK--VVALNKGAAHQAQRYALMVSTKDGQVATYATADAKAGLVKRT 762
DB 708 AVIEGNFSLRLKASDRVVVNNGAHKQAVRPLLLTTDNGIKAYHSQEA--AGLVRYT 765
```

RESULT 6

GTFC STRMU

ID GTFC STRMU STANDARD; PRT: 1455 AA.

AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;

DT 01-NOV-1988 (Rel. 02, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)

DE (Dextranase) (Sucrose 6-glucosyltransferase).

GN Name=gTfC; OrderedLocusNames=SMU.1005;

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

RN [1]

144 IPSDLAKNSNVKQDKGYYYYDQGNVKNFVSGDKIYYFDETGAYKDTSKVDKSS 203
178 IAA-LGNVKNIRKVGKYYYYKEDGTLQKYNALNGKTFDFDETALSNNT-LPSKGN 235
204 SAVSONATTIFAANNRAYSTSAKNEAVDNLYTADSNYRPSKSLKOGKWTESKDDFRL 263
236 ITNNQNTNSFAQYNQVYSTDAANFEHVDHYLTAESWYRPKYILKOGKWTQSTKDFRL 295
264 LMAMPDTEKKNYNNKVVIGIDKTYTAETSOADLTAAAEVLQARIQKITSNNNTKW 323
296 LMTWPDQETQRYVNNYNAQLGIHQTYNTATSPQLNLAAQTIQTKEIKTAEKYTNW 355
324 LREALSAFKTQPOWNGESEKYDDHLQNGALLFNQDTLTPDQTSNYLLNRTPTNQTG 383
356 LRQTISAFVKTSANWSSEKFPDDHLQKGLLYSNNKSLTSQANSNYRLNRTPTNQTG 415
384 SLDSTFTYNNPDPLGSGYDFLLANDVDNSNPVQAEQLNWLHYLLNFGSIYANDADANFDS 443
416 KQDPRT--ADRTIGGYEFLANDVDNSNPVQAEQLNWLHYLLNFGSIYANDADANFDS 473
444 IRVDAVNDVADLLQISSDYLKAAAYGIDKNNKANNHVSIVEAWSNDPTPYLHDDGDNL 503
474 IRVDAVNDVADLLQIAGDYLKAAKGIHKNDKANDHLSILEANSYNDTPYLHDDGDNI 533
504 NMDKFRSLMWSLAKPLDKRSLNPLIHNLSVDREVDREYETVPSYSFARAHSEVQD 563
534 NMDNRLSLSLYLAKPLNQRSGMNPITNSLVNRTDDNAETAAPVPSYFIRAHSEVQD 593
564 IIRDLIKAEINPNSGYSTQBEIEQAQKIYNEDLKTDKTKYTHNVPLSYLLITNKS 623
594 LIRNIRABINPNVVGYSYFTEIEIKKAEIYNKOLLATEKTYTHNTALSYALLITNKS 653
624 IPRVYGDMPFTDDGVMAKNTVNYDAIESLLKARKYVSGGQAMQNYQIENGELTIVRY 683
654 VPRVYGDMPFTDDGVMAKNTVNYDAIESLLKARKYVSGGQAMQNYQIENGELTIVRY 713
684 KGKALKQSDGDATTTRTSGVYVGMQNPNSLDGK-VVALNMGAAHQAQYRALMVSTKD 742
714 KGKALKATDGTGRTTSTSGVAIEGNNPSRLKASDRVNVNMGAAHQAQYRPLLTND 773
743 GVAITYATADASKAGLVDENGILYFLNDDLGKVPANPOVSGFLQVWPVGAADQDIR 802
774 GIKAYHSDQEA--AGLVRYTNDRGELIFTAADIKGYPANPQVSGYLGWVPVGAADQDVR 831
803 VAASDTASTDGKSLHODAMDSRVNPFEGSNFQSEATKEEYTNVVIANNVDKFSWGIT 862
832 VAASDTASTDGKSLHODAMDSRVNPFEGSNFQSEATKEEYTNVVIANNVDKFSWGIT 891
863 DFEMAPQVYSTDGSFLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLK 922
892 DFEMAPQVYSTDGSFLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLK 951
923 VMADVVPQMYTFPKQEVVTVTRTDKFKPIAGSQINHSILYTDYTKSGDDYQAKYGAF 982
952 VMADVVPQMYTFPKQEVVTVTRTDKFKPIAGSQINHSILYTDYTKSGDDYQAKYGAF 1011
983 LDELKEKYPFELTKQISTGQAI DPSVKIKQMSAKYVNGSNILGRGADVLSDOVSNKYF 1042
1012 LDELKEKYPFELTKQISTGQAI DPSVKIKQMSAKYVNGSNILGRGADVLSDOVSNKYF 1071
1043 NVASDTLFLPSLLG-----KVVEGIRYDGKGYIYNSSATQDVQKASFITBAGNLYYFG 1097
1072 SLVSDNTFLPSLVNPNHGTSSVTVGLVDFGCKGYIYST-SGNQAKNAFISLGNWYF 1130
1098 KGVYVTVGAQTINGANYFFLENGTALNTIYTDAGNSHYANDGKRYVYQOQGNDR 1157
1131 NKGVMVTGAQSINGANYFSLNGIQLRNAIYDNGNKVLSYVNDGRRYVYGLFQGW 1190
1158 YFGKGNMAVGLTVVDGNYQYFDKGVQAKDKLIVTRDGVYFDQHNGNAATNTFIADKT 1217
1191 YFQNGIMAVGLTRIHGAVYFDASGFQAKGQFITTADGKLYFDRSDGNGQISNRFVRSK 1250
1218 GHWYVLGKGVAVTGAQTVGKQLYFEANGQGVKGFVTSDEGKLYFYDVSDDMWDTTF 1277

RESULT 7

GTFD_STRMU

ID GTFD_STRMU STANDARD; PRT; 1462 AA.
AC P49331; O69383; O69386; O69389; O69392; O69398;
DT 01-FEB-1996 (Rel. 33, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-S precursor (SC 2.4.1.5) (GTF-S) (Dextranucrase)
GN Name=gtfD; OrderedLocNames=SMU.910;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RC MEDLINE=91100958; PubMed=2148600;
RA Honda O., Kato C., Kuramitsu H.K.;
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyltransferase-S enzyme";
RL J. Gen. Microbiol. 136:2099-2105(1990).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f, MT4467 / Serotype e, and MT8148 / Serotype c;
RC MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S., Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RC MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P., Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental pathogen";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 6 cell wall binding repeats.

```
QY 729 ANOEYRALMVSTKOGVATYATDADASKAGLVKRTDENGVLFLNDDLKGVANPOVSGFLQ 788
Db 765 KNOEYRPLLTGKTGTS:SDAAK--SLYRTDKDGBLFDASDIQGYLNPQVSGYLA 822
QY 789 VWPVGAADDQDIRVAASDTASTDGLSLHQDAAMDSRVMFEGSFNFQSFATKEEETNVV 848
Db 823 VWPVGAASNDQVRVAASKANATQGVYESSALDSQLIYEGSFNFQDFVTKDSYTNKK 882
QY 849 IANNVDKFKVSGITDEMAPOQVSVSDGFLSVIONGVAFTRDRYDLGMSKANKYGTADQ 908
Db 883 IAQNVQLFKVSGVTSPEMAPOVSVSDGFLSDIIQNGVAFEDRYDLGMSKANKYGSQD 942
QY 909 LVKAIKALHAKGLKYMADWVDMQTFPKQEVVTVTRTDKFGKTAGSINSLIYDTPK 968
Db 943 MINAKVALKSGIQVADWVDMQTFPKQEVVTVTRTDKFGKTAGSINSLIYDTPK 1002
QY 969 SSGDDYQAKYGGAFDELKEKYPELFTKKQISTGQAIDPSVVIKOWSAKYFNGSNILGR 1028
Db 1003 SNGKDYQAKYGGAFDELKEKYPELFTKKQISTGQAIDPSVVIKOWSAKYFNGSNILGR 1062
QY 1029 ADYVLSDQVSNKYFNVASDTLFLPSLLKVVESGIRYDGKGYIYNSSATGQVKAFTT 1088
Db 1063 VGYVLKDNASDKYFELKNGQTLPLKQMTNKEASTGFVNDGNGMTFYST-SGYOAKNSFVQ 1121
QY 1089 EA-GNLYYFGKDGVMYVGAQTINGANYPELENGTALRNTIYTDAGNSHYIANDGKRYEN 1147
Db 1122 DAGKNWYFDDNNGHMYVGLHNGEVQYFELSGVQLRESFLENADGSKNYFGLGNRYSN 1181
QY 1148 GYQOFEND--WRYF-KDGNMAGLTVVGNVGYFQKDGQVQAQDKIIVTRDGKRVYFDQIN 1204
Db 1182 GYSPDNDSKWYFDFASGVMAGLKTINGTQYFDQDGQVQKAGNITGSDGKRYFDDGS 1241
QY 1205 GNAANTFTIADTGHYVYLGKDGVAVTGAQTVGKQKLYFEANGQVQKDGFTVSDGKLYF 1264
Db 1242 GNMAVNRFPANDKGDWYVYVLSNGIALVGVQTINGKTYFVGQDGKQIKGIIT-DNGKLY 1300
QY 1265 YVDVSGDMWTFIEDKAGNFYLGKDGVAVTGAQTVGKQKLYFEANGQVQKDGFTVSDGKLY 1324
Db 1301 FLANSGLARNIIFATDSQNNWYFSGDGVAVTGSQTIAKKLYFASDGKQVKSFTV-YN 1359
QY 1325 GKIRYDADSGEVOFNKTVKAADGKTYVIGNDGVAVDPSVVKGTQFKDASGALRFVNLKG 1384
Db 1360 GKHYHADSGLQVNRFEADKDG----- 1383
QY 1385 QLVTSQGWYETANHDWYIQS--GKALTGEQTINGQHLFKEDGHQVQKQOLVTGTDGKRY 1443
Db 1384 -----NWYVLDNSGEALTSQQRINGQVRFVFTREGKQVKGDVAVDERGLARY 1429
QY 1444 YDANSQDQAFNKSVTIV-NGK 1462
Db 1430 YDRNSGNMVKVVTIANGR 1449
```

RESULT 8

```
Q541178_STRGN PRELIMINARY; PRT; 1577 AA.
ID Q541178; Q54247;
AC 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Glucosyltransferase.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
  phase variants.";
```

```
RL Dev. Biol. Stand. 85:309-314(1995).
RP [2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RA Minick P., Vickerman M.;
RL Submitted, (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U12643; AAC43483.1; -; Genomic_DNA.
DR FIR; B41898; B41898.
DR HSSP; P06653; 1HCX.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AB0328DC5E08D18 CRC64;

Query Match 43.2%; Score 3604; DB 2; Length 1577;
Best Local Similarity 44.9%; Pred. No. 6e-158;
Matches 756; Conservative 252; Mismatches 457; Indels 220; Gaps 35;

QY 1 MEKNVRFKMHKVKKEWVTLVSASATMLAS--ALG-----ASVASA 38
Db 2 MEKNVHYKMHKVKKEWVTLVSALLVAPKALGLESVYADDANQVTVNVEQSASQSK 61
QY 39 D-----TDTASDSSN-----QAVVTGDTNNQATDQTSIAATA----- 72
Db 62 DSEQTTSKATDSSQLEKVEQASSSKETYQASAAATPANEQTQQDKEVETSRDTSRHE 121
QY 73 -----TSQSASTD--AATDQASAAEQTGT--TASTDTAA-----QTTNANEA-- 113
Db 122 LTQKTSDDSSKSSGSSQEPKVADQAESTDKTQAALQAKQDSRANDQEETENVAKATSD 181
QY 114 KWPVTENE-----NQGFDEML---AEAKNVATAESDISPSDLAKMSNVKQVDGKY 161
Db 182 KIATPPKKERLPEPAQRKESITERKMLAAQAAPVNTHEHDDV-----LAHKTIDGKK 235
QY 162 YYVDQGNVKNKFAVSGDKIYFDF-ETGAYKDTSKVDADKSSASVSONATTI PAANNRAY 220
Db 236 YVVDQDGTVKNFAVELNGKLYLFDATGALVDSNEYQFQGTSSINNE---FTQKNAY 292
QY 221 STSAKNFAVNYLTADSWYRPSILKDGKTWESGKDDFRPLLMAWMPDPTETKRYNY 280
Db 293 GTTKDIEVTDGYLTADSWYRPKFILKDGKTWASTETDLRPLLMAWMPDKRTQINLYN 352
QY 281 MNKV-VGIDKTYTAETSQADLTAAELVQARIEQKITSENNTKWLREISAFTVKTPQWN 339
Db 353 MNQENLGIG-APESKTEQVLLTNVAVQVQVQRIEERISKEGDTKWLRLTMSAFVKTQPNW 411
QY 340 ---GESEKPYDDHLQNGALLFDNOTDLTPDTQSNRYLLNRTPTNQTGSLDSRFTYND 395
Db 412 IKTSETTGTNKHLOGGALLYTN-SDKTSANSRYLLNRTPTTSQTGT-PKYFIDKSN- 468
QY 396 PLGGYDFLLANDVDNSNPVQAEQLNWLHYLLNFGSIYANDADANFDSIRVDADVNDAD 455
Db 469 --GGYEFLLANDFDNSNPAVQAEQLNWLHYLWVNFSGIIVANDPTANFDGVRVDADVNDAD 526
QY 456 LLIQSSDYLKAAYGIDKKNKANNHVSIVEAWSNDTPLYLHDDGNLMMNMKFLSLMLW 515
Db 527 LLIQASDYFKSRYKVGSEEEALKHLSILEAWSNDPDYNDKDTKGAQLAIDNKLKSLLY 586
QY 516 SLAKPLDKRSGNLPLHNSLVDRVDDREVTVPSPYSFARAHDSVQDIIRDIKAENP 575
Db 587 SFMRKLSIRSGVEFTIYNSLNDSTRSTENKNGERTANYIFRAHDSVQTVIADIIRKNP 646
QY 576 NSFGYSFTQEEIEQAFKIYNEDLKKTKYTHYNVPLSTLLLTNKGSIPIRYVYDGMFTD 635
Db 647 NTDGLTFTMDLQKAFKIYNEDMRKADKKYQFNIPTAHALMLSNKDSITRIVYDGLYTD 706
QY 636 DQGMANKTVNYDAIESLLKARKMYVSGQAMQNYQIG-----NGEILTSVRVYQK 685
Db 707 DQGMKKSYPYHDAIDALLRARIKYVAGQDMKVTYMGVPPREADKWSYNG-ILTSVRVGT 765
```

```
QY 686 GALKQSDGXDATRTSGVGVNGNQPNSL-DGKVALNMGAHAHQYRALMVSTKGV 744
Db 766 GANEATDEGTAEATRTQGMVAISNNPNLKLNEWDKLQVNMGAHKQYRVPVLLTKDGI 825
QY 745 ATYATDADAKAGLVKRTDENGILYFLNDDLKGVANPOVSGFLOVWVPVGAADDIRVA 804
Db 826 SRYLTDEEVPQS-LWKKTIDANGILFDWMDTAGYNNVQVSGYLAWVVPVGAKENODART 884
QY 805 ASDTAQTGKSLHQDAAMDSRVMFEGSFNFQSFATKEBEYTNVTVANNVDKFSVGTDF 864
Db 885 ASKKKNASGVYESSPALDSQLYEGSFNFQDFATRDQYTNKVIAKNVLFKEGVTSF 944
QY 865 EMAPYVSTGQFIDSLVQNGYATDRYDLGMSKANKYGTADQLVKAIKALHAKGLKM 924
Db 945 ELPYVSSQDCTFLDSITQNGYAPEDRYDMAMSKNNKYGSLDDLNALRALHSHVNIQAI 1004
QY 925 ADMVPDQMYTPKQEVVTVTRTDKFGKPIAGSOINHSLYVTDTKSSGDDYQAKYGAFLD 984
Db 1005 ADMVPDQYINLPKKEVTVATRVNNYGYRBEAGIEKELYVANSKNTGTDYQKYGAFLD 1064
QY 985 ELKEXPELFTKKQISQGAIDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNV 1044
Db 1065 ELKAKYPELFEVQISNGQMTTDEKITKSAKYFNGTNILGRGAYVYLKOWGSKYLSN 1124
QY 1045 ASDTLFLPSSLLGKVVESGIRVDGKYIYNSATGDQVKAQFI-TEAGNLVYFGKGYMV 1103
Db 1125 KNGETALPKQLVKEASTGFVKDITNGFKFYST-SGNAQKDTPIQDENGWYVFDNQGLV 1183
QY 1104 TGAQTINGANYFLENGTALRTIYTDAGNSHYANDGKRYENG-YQFQNDWRYF-KD 1161
Db 1184 TGAREIDGKQLYFMKNGVQLRDALQEDENGQYVDKTKGAKVNLRYTSDGQNWRYFQAK 1243
QY 1162 GNMVGLTVVDCNVGYFQKDGVOAKDKIIVTRDGKRVYFDHNGNAATNFI-ADKTGHW 1220
Db 1244 GWMAGLVKIGDQOYFDONGYQVKGVVRAKDGKLYRFDKSGNAVLRNFAQGNPDSW 1303
QY 1221 YLKGKGVAVTGAQTVGKQKLYFEANGQOVKGFVTSDEGKLYFYVDVSGDMWTDPTFIED 1280
Db 1304 YFYGADGVALTGLQIGQOTLYFGQDGKQVGVVLMADKSIYRFDANSSEMVAVKFAEG 1363
QY 1281 KAGNWFYLGKGAATGAQTIRGKLYFKANCOOVKGDIVKGTGKIRYDAKSGEOVFN 1340
Db 1364 AKNEWYFDQGKAVTGLKTIINNQLYFDQDGKQVGVVTLADKSIYRFDANSSEMVAV 1423
QY 1341 KTVKAADGKTYVIGNDGVAVDPVSVVKGQTFKDGASGALRYNLKGLVTSQWYETANHDW 1400
Db 1424 K-----FAEGAKNEW 1433
QY 1401 VYI-QSGKALTGEQTINGOHLVFKEDGHQVKGQLVGTGDKRVYVDANSQDQAFKSVTV 1459
Db 1434 YFQDQGRAVTGLQTIINQLVYFGQDGKQVGVVTLADKSIYRFDANSSEMVAVK----- 1489
QY 1460 NGKTYFQNDGTAQTAGNPKGQTFKDGSDIRFYSNEGQVLTGSGWYENAGQWLYV-KNG 1518
Db 1490 -----FAEGAKNEWYFDQDG 1505
QY 1519 KVLTLQTVGSORVYFDENGIOAKGKAVRTSDGKIRYFDANSSEMITNOW-KFVVGQVYV 1577
Db 1506 KAVTGLKTIINQLVYFGQDGKQVGVVY-DGAEYFDPKSGDMVRNKNWIRLEDGTWY 1564
QY 1578 FQNDG 1582
Db 1565 FDRNG 1569
```

RESULT 9

```
Q9LCH3 STOR
ID Q9LCH3 STOR PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

```
DE Glucosyltransferase.
GN Name=gtfR;
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=11303;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC10557;
RX DOI=10.1128/JAI.68.5.2475-2483.2000;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RL encoding glucosyltransferase from Streptococcus oralis.";
DR EMBL; AB025228; BAA95201.1; -; Genomic_DNA.
DR HSSP; P06653; 1HCX.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding 1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26B4D7C2E543 CRC64;
```

```
Query Match 42.2%; Score 3527; DB 2; Length 1575;
Best Local Similarity 44.0%; Pred. No. 2.2e-154;
Matches 742; Conservative 250; Mismatches 482; Indels 212; Gaps 29;
```

```
QY 1 MEKVRFPKHVKKGRWTLVSASATMLAS-----ALGASVASAD----- 39
Db 2 MEKHIYKMHKVKKNWVAIALTTLALIVAPKVLGLESVGVHADDVKVWVQSPATAQTS 61
QY 40 -----TDTASDDSNQAVVTGQTT-----NNOATDQTSIAATATSEOS 77
Db 62 PQQTPAQAKIASQEAQKVTADKVTDDVAASEKPAKPAENTATVQTN--AQEPAPKA 119
QY 78 ASTDAATDQASAAEQTGTTASTTAQTATTNANEAKWVPTTENENQGTDEMLA-EAKNV 136
Db 120 DTKEASTEKAASAEVKAANAITEIPKTEVADQNKQARPTTAQDQEGDKREKTA 179
QY 137 A---TASDSITPSDLAK-----MSNVKQVQDKYIYYQD 167
Db 180 ANPKVAKKORLPFEPGSKQGAIAERMAVDAQAPAPVADHDDVLSHIITDGNYYVQDD 239
QY 168 GNVKKNFAVSGDKIYYFD-ETGAVKQTSKYVDADKSSASVQNAITIPANNRAYSTSAKN 226
Db 240 GTVKKNFAVELNGRILFPDAETGALVDSNEYVQFGQTSSLNNE---FSQKNAFYGTTDK 296
QY 227 FEAVDNYLTADSWYRPKSILKDGKWTBESGKDDFRPLLMAMPDTETKRNYYNMYNKKV 286
Db 297 IETVDGYLTADSWYRPRKFLKDGKWTASTETDLRLPLLMAMPDKRTQINVLNMYNQGL 356
QY 287 IDKTYTATSDADLTAAAEVQARIEQKITSNNTKMLREASAFVKTPQOWN-----GES 342
Db 357 GAGAFENKVEQALLTGASQQVQRKIEEKIGEGDTKWLRTLMGAFVKTPQNMNIKTES 416
QY 343 EKPVDHDLQNCALLFDNQTDLTPTQSNRYRLNLTPTNOTGSLDSRFTYNNPDUGGYDF 402
Db 417 TGTKKDHLQGGALLYTN-NEKSPHADSFKRLNLTPTPSQTGT-PKYFTDKSN---GGYEF 471
QY 403 LLANDVNSNFPVQAEQLNWLHLLNFGSIYANDADANFDSIRVDAVDNVDADLLQISDD 462
Db 472 LLANDFNSNFAVQAEQLNWLHLMNFGSIVANDPTANFDGVRVDAVDNVDADLLQIASD 531
QY 463 YLKAAYGIDKKNKNNHVSIVEAWSNDTPLYLHDDGDNLMNMKNKFLSLMLSLAKPLD 522
Db 532 YFKSRYKVGSESEEAIKHLSILEAWSNDPPYNNKDTKGAQLAIDKNKLSLLYSMPRLS 591
QY 523 KRSGLNPLHNSLVDRVEDDREVEVTPSPYSFARAHDSQVQDIIRDIKAEINPNPSFGYSF 582
Db 592 IIRSGVEPTITNSLDRSEKKNKNGERMANYIVRAHDSEVQIVADIIRININPNTDGLTF 651
```


Db 588 TDGFTTLDQLKQAFDINYNKDMNSVEKHYYTHYNI PAAYAVMLSNMESVTRVYVYGLFTDD 647
QY 637 GQYMAKTVNYDAISLLKARKMYVSGGQAMQNYQ-----IGNGEILT 679
Db 648 GQYMETKSPYDAINTLLRARIYAAGGTMEHKAITYPSAAMKAKNPDSGSLVGNSEVLV 707
QY 680 SVRYGKALKQSD-KGDATTRTSVGVVGNQPNFSLD-GKVVALNMGAAHANQEVRLM 737
Db 708 SVRFQGVMSADMTGQGLAKTSGMFSLIANNPELELDANEIEIKVNVGKIHAQAYRPLL 767
QY 738 VSTKQGVATYATDADASKAGLVKRTDENGLYFLNDLKGVANPOVSGFLQVWVPVGAAD 797
Db 768 LTTDKGLQKLYNDSNTN---LTKVADKQGFIFKGEIKGYKQVEVNGVLSVWVPVGAKS 824
QY 798 DQDIRVAASDITASTG-KSLHQAADWSRMVEGFSNFOSPATKBEETVNVVIANVVKF 856
Db 825 DQDIRVAASKANAKGDKSYTASQALDSQLIIEGFSNFQDFQKDAQYTNKKIAENTOLF 884
QY 857 VSWGILDPFEMAPQVYSSTDGFLDSVIONGYAFTDRYDLGMSKANKYGTADOLVKAIKAL 916
Db 885 KAWGVTSPFMAPQYVSATDGTFLDSIIQNGYAFSDRYDLAMSKNKYGSKEDLANALKAL 944
QY 917 HAKGLKMWADWPDOMYTPPKQEVTVTRTDKFGKPIAGSQINHSILYVTDKSSGDDYQA 976
Db 945 HAAGQAIADWPDIQIYQLPGKEVTVASRVNRYGRVKIDQPNVKLYLANTHSSGKDFQA 1004
QY 977 KYGGAFLDELKBYPELFTKQISTGQAIIDPSVKIKQMSAKYFNGSNILGRGADYVLSGD 1036
Db 1005 KYGGEFLAELQXQPEMFTAKMISTGKPIDSSVLKQWESAQYFNGTGNVLGRGTDVLSDE 1064
QY 1037 VSNKYPNVAADTLFLPSSLLG-KVVEGIRYDGKGYIYNSSATGQVQKASFTIEAGNLYY 1095
Db 1065 GTGKYFTVNERGEFLPAVLGTGKEAKTGFYNDGKMTYFTTA-GSQAKSDFTVAGNTYY 1123
QY 1096 FKGQGVMTGAOTINGANVFFLENGTALJNTIYTDAGNSHYVYANDGKRYENGYQOFGND 1155
Db 1124 FDTGHWITGPGINTKFFYFLPNGVMLKDAWEDDRGSRVYVYKGTVMYKGSR---NNE 1180
QY 1156 W-----RYFKD-GNMAVGLTTVDGNVQYFDKGVQAKDKIIIVTRDGKRYFDQ 1202
Db 1181 WFAWTDKQGLRFRHFDNYGFMVGLVTHGNVQYVDEBGFQVKGDFVTDKAGQTRYFDK 1240
QY 1203 HNGNAATNTFIADKTHGYVYLGKQGVAVTGAOTVQKQKLYFEA-NGQOVKQGFVTSDEGK 1261
Db 1241 NTGNLVKGOFF-NQNGHWIYSDQGLIAKGAQTIKQKLYFPAKTGAQVKGDFVTDKGN 1299
QY 1262 LVPYDVDSGDMWTDTFIEDKAGNWFYLGKQGAATVGAOTIRGOKLYFKA-NGQOVKQGDIV 1320
Db 1300 TFFYSDDTGLAVSTFFSTGNNAWVADENGHVAKEKTINGOKLYFDTKTGQQAQGRPV 1359
QY 1321 KGTDGKIRYDAKSQEQVFNK--TVKAADGKTYVIGNDGVAVDPVYVKGQTFKDAAGALR 1378
Db 1360 RDAKG-LRFYDADTGAULTNSFLETKAGSNQWYMGADGYAV-----1400
QY 1379 FYNLKGQLVGSGWETANHDWYIQSCKALTGEBQTINGQHLVFE-KEDGHQVKGQLVTCGT 1437
Db 1401 -----RGHQTISQRHMYFDAETGQQAQGVVTD 1429
QY 1438 DGKVRYYDANSQDAFNKSVTVNGKTYVFGNDGTATAGNPKGQTFKQSDIRFRYSMEQ 1497
Db 1430 NGRKYPYDANTGDRVNVQVLVNGSWYFFGYQAA-----1464
QY 1498 LVTGSGWYENAOQMLYVKNQKVLTLGLQTVGSQRVYFDENGIQAKGKAVRTSDGKIRYVD 1557
Db 1465 -----VTGFRDIRGQHLVFNPDGTQAGTGVTKI-DNRIYTFD 1500
QY 1558 ENSGSM 1563
Db 1501 ADSGEL 1506

RESULT 13
Q56CX8_9STRE

ID Q56CX8_9STRE PRELIMINARY; PRT; 1506 AA.
AC Q56CX8;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Glucosyltransferase-T.
GN Name=gtft;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI TaxID=1310;
RN 11)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
KW EMBL; AY966490; AAX76986.1; -; Genomic_DNA.
SQ SEQUENCE 1506 AA; 167016 MW; 94B882BF2C17C451 CRC64;
Query Match 40.8%; Score 3410.5; DB 2; Length 1506;
Best Local Similarity 43.4%; Pred. No. 5e-149;
Matches 708; Conservative 277; Mismatches 446; Indels 197; Gaps 39;
QY 1 MEKNVRFQWHKVKKRWVTLVSASATMLASALGASVASADTDTASDD--SNQAVVTGDQTT 58
Db 1 MERKLHYKLHKVKQWVTVIAVASAG--LASIVGA--GSLQTVSADDLAKEQAAASQQA 57
QY 59 NNQATDQT-----SIAATATSE-----QSASTDAAT-----DQASAAEQTGTTASTDT 102
Db 58 ANQNEDEVASDAADTASAKATSEKVVQSSDTNSETNQVETKQOASAKESADAVAKQAPQ 117
QY 103 AAQTTN---ANEAKWVPTENENQGTDEMLAEAKNVATAESDSIPSLAKMNVKQVDG 159
Db 118 AGPATTSQVASSESSVAPSKB---ADK--AAAGSVSQNEBEAALS---LANIKKIDG 167
QY 160 KYIYYDQGNVKNKFAVSVGDKIYYFD-ETGAYKDTSKVDADKSSSAVSQNAITFAANNR 218
Db 168 KYIYYMADGSKYKONFAITVQGMLYFDKTAGLSSTYSFSQGLTPIVSD---FSVNNK 224
QY 219 AYSTAKNFEAVDNYLTADSWVRPKSILKDGKWTPESGKDDFRPLLMAWMPDTETKRYV 278
Db 225 AFDSEKSEFELVDGLTAESWVRPAKILENGKTWVDSKETDLRPVLSWPNKDTQVAYL 284
QY 279 NYMNKVIGDKTYTAETSQADLTAAELVQARIEOKITSENNTKWLREAI SAFVKTQVQW 338
Db 285 NYMSKALGKGEFTTETSQTLTAELIQAIEARVSKEQGTWLRMAAFVATQSRW 344
QY 339 NGESEKPYD--DHLONGALLFDNQDTLTPDTQSNVRLNRTPTNQTGSLDSRFTYNPDP 396
Db 345 NKDSQ-YDKADHLQGGALLYTN--NNLTWANSNRLNRTPTRO-----DGKTHYSKADK 398
QY 397 LGGVDFLLANDVDSNPVQAEQLNWLHYLNFSGIYANDADANFDSIRVDAVDNVDADL 456
Db 399 YGGVEFLANDVDSNPVQAEMLNQIHYLNMWGEIVMGDKNANFDGIRVDAVDNVDAT 458
QY 457 LQISDYLKAAAYGIDKNNKNANNHVSIVEAMSDNDTPYLHDDGDMNMNDKFRLSMLMS 516
Db 459 LQLYTNFNSVGVNKSQAQALAHISVLEAWSYNDNDYNTQDTNGAALAMDNGLRFSLLY 518
QY 517 LAKPLDKES-GLNPLIHN--SLVDRVDDREVEVTPSYSFARAHDSVQDIIIRDIKAEI 573
Db 519 LTRPINETPGMSTLIKSEYGLTDRTKNDKYGDTQPSYFVFAHDSVQTVTAQIIKEKI 578
QY 574 NPNSFGYFTOBEIEQAFKIYNEIDLKTKDYTHYNVPLSYTLTLTKSGSI PRVYVGMF 633
Db 579 DPTDGTFTLDQLKQAFELYNKONSVNKHYYTHYNI PAAYAVMLSNMESVTRVYVYGLD 638
QY 634 TDDGQYMAKTVNYDAISLLKARKMYVSGGQAMO--NYQ-----IGNGE 676
Db 639 TDDGQYMAKSPYYDAINTLLRARIRYAAGGQIMEHNSYKPSAAMKAHPADAGNVLGSE 698
QY 677 ILTSVRYGKALKQSD-KGDATTRTSVGVVGNQPNFSLD-GKVVALNMGAAHANQEVYR 734

Search completed: February 11, 2006, 19:38:29
Job time : 223.546 secs

589 QAPKIYNEDLKTKKXTHNVPLSYTLTNTKGSIPRVYVYDMFTDDGQYMANKTWYD 648
669 QAFYIYNADDELKADKEYATYNI PASYAVLLTNKDTVPVYVYDGLFSDGQYMSQKSPYD 728
649 AIESLLKARMKYVSGGOM-----QNYQINGGE-----ILTSVRVKGALKQSDKGDAT 697
729 AITSLKSRKXYVAGQSMNMYLHCFDPAPKNETKPPQGVLTSVRYGKAMTADDLGNSD 788
698 TRTSGVVMGNQPNFSL-DGKVVALNMGAAHAHQYRALMVSTKDG VATYATDADASKA 756
789 TROGGIGLVINNKPFNLNDDDEIVLNMGAHKNQAYRPLMLTTKSGLOIYDKDAGAP-- 846
757 GLVTRTDENGILYFLNDDILKGVANPQVSGFLOVWVPVGAADQDQIRVAASDTASTDGSKL 816
847 --VVTYNDAGQLIFKSDMVYGSNPQVSGYFAAWVPVGASDSQDARTQSSQSETKDGDDV 904
817 HODAAAMDSEVMEFSGFNQSFATKEEYTNVIVIANNVDFVSVWGIITDFEMAPQYVSSDVG 876
905 HSNAAALDSNVIYEGFSNFQAMPEKNDFFTNNKIAQNAKLFKOLGITSFELAPQYRSSTDN 964
877 QFLDSVIQNGYAFTRDYDLGMSKANKYGTADOLVKAIKALHAKGLKVMADWVPDQMYTFP 936
965 SFLDSVIQNGYAFTRDYGVYNTPTKYGTVDQLLSLRAHQAQIOAINDMWVPDQIYNLP 1024
937 KOEVTVTTRDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDELKEKYPELFTK 996
1025 GEQIVTAVRTNGSGKYDYDSVINNTLYDSRTVGGG-EYQKFGGLFLDQKXDYPSLPFET 1083
997 KOISTQOAITDPSVKIKQWAKYFNGSNILGRGADYVLSQVSNKYFNVASDTLFLPSSLJ 1056
1084 KOISTQNPQNPVVKIKWAKYFNGSNIOGRGAWYVLDKQWATNQYFNVSNDGFLPKQLT 1143
1057 GKWBESG-IRYDGKGIYNSSATGDOVKASFTEAGNLVYFGKGVMTGAOTINGANYF 1115
1144 GEKSTGFTENGKTSFYSTS--GYQAKDTFTQDGTNWTYFVFNAGVMTGKQNIHDKNYI 1201
1116 FLENGTALNTIYTDAGNSHYIYAND-GKRYENG-YQFQNDWRY-FKDGNMAGLTVV- 1171
1202 FLPNGVELQDAYLFD-GNOEFYVYKAGEQVMNQYQDSQNMWHYFFENGRAIGLITEVP 1259
1172 --DGN--VQYFPGKDGVOAKDKLIVTRDGKRVYFDQHNGNAATNTP--IADKTHWYILGK 1225
1260 NADGTHVTQYFDANGVQIKGTAIKQNNQLRYFDEATGNMVVNSWGLADKS--WLYLNA 1317
1226 DGVAVTGAQTVCKQKLYPEANGQVKGDFTVSDGKLYFYDVYDVGDMWTDTFIEDKAGNW 1285
1318 QGVAVTGNQKIDGEEYFNADGKQVKGNAIIDNNGDQRYIDGDKGVVNSWGLPDSGW 1377
1286 FYLGKDGAAVTGAQITRGQKLYFKANGQVKGDI VKGTDGKIRYIDAKSGEQVFNKTVKA 1345
1378 LYLNKGIATVGRQVINNVQNFVFGNDGKQIKDAFKLLSDGSMVYLLDDKG--LITTGAKV 1434
1346 ADGKTYVIGNDGVAVDPSVVKQOTFKDASGALRFNLK-GQLVTSQWYETANHDWYI- 1403
1435 INGLMFFDKDG-----HQIKGDASTDANGKRYIDKNDGHLVTNS-WGELPDGSLWYLE 1488
1404 QSGKALTCBOTINGQHLVFKEDGHQVKGOLVTGTG-----KURY 1443
1489 EOGDAVTGQVYDGTTRYFDEDEGKQIKNSLKTLANGDKIYLDGDGVAATGLQHVGDKIMY 1548
1444 YDANSQDAFNKSVTV-NGKTYVYFNGDGTQATA-GNPKQOTF-----KDGSDIR----- 1490
1549 FD-EDGKQVVGKFSVAKDGSWYLLNQDGVAAVGPSSINGQSLYFDQDGKQVKNVNRNSD 1607
1491 -----FYSMEGQVLTGSGMYENAOQOMLYV-KNGKVLTLGLQTVGSRVYFDENGIQAKGK 1544
1608 GTTNYTYTGLTGEKLT-QDFGELPDGSWIYLDQAQGHVTVTGAQIINGQNLVFKADGQOVKGH 1666
1545 AVRTSDGKIRYFDENSGSNITNQWKFVY-GQYVYFNGDGARI 1585
1667 AYTDLGHMRFYDPSGDMLSNRFEQITPGVWAYFADGVAI 1708

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 154.681 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-38
Perfect score: 8237
Sequence: 1 MEKKLHYLKHVKYKHWITIA.....FVRIGDQWYFGNGAATNL 1554

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8237	100.0	1554	7 ADD93658	Add93658 Streptococcus
2	8237	100.0	1554	9 ADX37281	Adx37281 Streptococcus
3	3818	46.4	1518	7 ADD93660	Add93660 Streptococcus
4	3818	46.4	1518	9 ADX37283	Adx37283 Streptococcus
5	3615	43.9	1430	5 AAU98029	Aau98029 S. mutans
6	3615	43.9	1430	7 ADD93656	Add93656 Streptococcus
7	3615	43.9	1430	9 ADX37279	Adx37279 Streptococcus
8	3610	43.8	1430	5 AAU98043	Aau98043 S. mutans
9	3609	43.8	1430	5 AAU98042	Aau98042 S. mutans
10	3609	43.8	1430	5 AAU98041	Aau98041 S. mutans
11	3604	43.8	1430	5 AAU98044	Aau98044 S. mutans
12	3604	43.8	1430	5 AAU98045	Aau98045 S. mutans
13	3502.5	42.5	1590	9 ADD93657	Add93657 Streptococcus
14	3502.5	42.5	1590	7 ADX37280	Adx37280 Streptococcus
15	3497	42.5	1476	5 AAU9284	Aau9284 Streptococcus
16	3472.5	42.2	1592	2 AAR32925	Aar32925 Glucosyltransferase
17	3446.5	41.8	1475	5 AAU98036	Aau98036 S. mutans
18	3445.5	41.8	1475	5 AAU98035	Aau98035 S. mutans
19	3445.5	41.8	1475	5 AAU98034	Aau98034 S. mutans
20	3442.5	41.8	1475	5 AAU98038	Aau98038 S. mutans
21	3440.5	41.8	1475	5 AAU98037	Aau98037 S. mutans
22	3440.5	41.8	1475	5 AAU98032	Aau98032 S. mutans
23	3439.5	41.8	1475	5 AAU98031	Aau98031 S. mutans
24	3435.5	41.7	1475	5 AAU98030	Aau98030 S. mutans

25	3434.5	41.7	1475	5 AAU98027	Aau98027 S. mutans
26	3434.5	41.7	1475	7 ADD93654	Add93654 Streptococcus
27	3434.5	41.7	1475	9 ADX37277	Adx37277 Streptococcus
28	3430.5	41.6	1475	5 AAU98040	Aau98040 S. mutans
29	3429.5	41.6	1577	2 AAR91047	Aar91047 Alpha-D-galactosidase
30	3428.5	41.6	1475	5 AAU98033	Aau98033 S. mutans
31	3407.5	41.4	1475	5 AAU98039	Aau98039 S. mutans
32	3255	39.5	1499	7 ADC54806	Adc54806 Protein S
33	3161.5	38.4	1375	5 AAU98028	Aau98028 S. mutans
34	3161.5	38.4	1375	5 AAU79288	Aau79288 Streptococcus
35	3161.5	38.4	1375	7 ADD93655	Add93655 Streptococcus
36	3161.5	38.4	1375	9 ADX37278	Adx37278 Streptococcus
37	3142	38.1	2835	5 ABR98574	Abbr98574 Dextranase
38	3142	38.1	2835	6 ABR55594	Abbr55594 Amino acid oxidase
39	3080.5	37.4	1365	7 ADD93659	Add93659 Streptococcus
40	3080.5	37.4	1365	9 ADX37282	Adx37282 Streptococcus
41	3041	36.9	1527	7 ADC54807	Adc54807 Leuconostoc
42	3015	36.6	1527	5 AAU80055	Aau80055 Leuconostoc
43	2958	35.9	1477	9 ADY72696	Ady72696 Mutant de
44	2953	35.9	1477	9 ADY72733	Ady72733 Mutant de
45	2948	35.8	1477	9 ADY72732	Ady72732 Mutant de

ALIGNMENTS

RESULT 1
ADD93658
ID ADD93658 standard; protein; 1554 AA.

AC ADD93658;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus sobrinus glucosyltransferase-U.

KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.

PN WO2003075845-A2.

PD 18-SEP-2003.

PF 07-MAR-2003; 2003WO-US006962.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

PI Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

PT Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

XX Claim 16; Page 15; 49pp; English.

CC The present sequence is the protein sequence of Streptococcus sobrinus glucosyltransferase-U (GTF-U). Peptide fragments of GTF-U, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

[illegible]

PR 07-MAR-2002; 2002US-0362209P.
PR 08-AUG-2002; 2002US-0402483P.
XX (FORS-) FORSYTH INST.
XX Smith DJ, Taubman WA;
XX WPI; 2003-845091/78.
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 16; Page 16; 49pp; English.
XX
XX The present sequence is the protein sequence of Streptococcus salivarius
CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutants glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Diepitopic or
CC multi-epitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
XX Sequence 1518 AA;
Query Match 46.4%; Score 3818; DB 7; Length 1518;
Best Local Similarity 50.1%; Pred. No. 1.6e-193;
Matches 783; Conservative 215; Mismatch 434; Indels 130; Gaps 24;
QY 1 MEKLIHYKLHKVKKHWTTIAVASIGLVLVG-----AGTVSA---EDKVANDTTAAQTAVG 52
DB 1 MENKLIHYKLHKVKKHWTTIAVASVALATVIGLGLVTTSSVSADETFQDKTVTOSNGTTAS 60
QY 53 VDTGQDQATNDANTNTDITDADOSANTNOPAGSDOSNNQDQAKQDPTANTDRNOADS 112
DB 61 LVTSPEATKEADKRTNKADVLTPAKETNAVETAT-TTNTQATAEAATTATTADVA-VA 118
QY 113 QTDNNQATQATSPATDGTSGVORRA---ANVATAADQEGQTAPSEQKSAISLDNVKL 169
DB 119 AVPENKEAVVTTDAPAVTTEKAEQEPATVKAEEVNTVEVKAPEALKDSEVEAALSLEKTN 178
QY 170 IDGKYVYVQADGSYKKNFAITVNGOMLYPDSYDSTGALSTSTYSFSQGTNLNVDVDFSSHNK 229
DB 179 IDGKYVYVNEDESHKENFAITVNGQLLYPGKD-GALTSSTYSFTPGTNIYVDFGSINNR 237
QY 230 AYDSTAKSPELVNGYLITANSWYRPAIGILRNGQTWEASNENDLRPVLMSWMPDKDTQVAVY 289
DB 238 AYDSSEASPELIDGVLTAWSYRPASIIKDGVTWQASTAEDFRLPMLMAMPNVVDVQVYL 297
QY 290 NYMKVLSANEVTEVNTSVDNLKEAQSIQTKIEOKITSNDSNTOWLRTEAMEAFVAAQPK 349
DB 298 NYMKSVFNL-DAKYSTDKQETLRVAAKDIIQIEKIQAEKSTQWLRTEISAFVKTQPK 356
QY 350 WNMSTENFNGK---DHLQGGALLYTNSDLTPWANSDYRLNLTPTQDDCT--KKYFTEGG 404
DB 357 MNKETNSYKGGEDHLOGGALLYVNDSTPWNSTDYRLNLTATNTQGTIDKSLIDEQS 416
QY 405 E-----GGYEFLLSNVDNSNPVQAEQLNQLHLMNMGDIVMGDKDANFDGVRVDAVDNV 460
DB 417 DPNHNGGDFLLANDVLSNPVQAEQLNQIHYLMNWSGI VMGDKDANFDGIRVDAVDNV 476
QY 461 NADLLQVYSNYFKDNYKVTDSANALAHISILEASLNDNQVNDTNGTALSIDNSSRLT 520
DB 477 DADMLQLYNTYFREYGVNKSANALAHISVLEAWSLNDNHYNDKTDGAALAMENKQRLA 536
QY 521 SLAVLTK-----QBQRIDLSNLSSESVNKERAND-----T 551
DB 537 LLFSLAKPTKERTPAVSPLYNNTFTTQREKTDWINKDGSKAYNEDGTIVKOSTIGKYNE 596

QY 552 AYGDITPTYSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFLDQLDKAFKINVEDMAKNV 611
DB 597 KYGDASGNVVFIRAHNNVQDIIIAEIIKKEINPKSDGFTITDAEMKQFAEYKNDMLSSD 656
QY 612 KTYTHYNIPAAVALLLSNMESVPRVYVYGDLYTDDQYMAKKSPPYDATALMLQGRYAVYS 671
DB 657 KKYTLNIPAAVAVMLQNMETITRVYVYGDLYTDDGHYMETKSPYVDITVNLKMSIKYVS 716
QY 672 GQSQSEVH-----KYVGNNOILSSVRYQDMLMSADDTQGTDLSTSGTLVTLV 718
DB 717 GQQAQRSVWLPDGTGMDNSDVELYRTNEVYTSVRYKDIMTANDTEGSKYSRTSGQVTLV 776
QY 719 SNDPNLGLGDS-LTVNMGRAHQAQYRPLILGTQGVQSVLYKDS-----TNIVKYTDANG 774
DB 777 ANPNKLNLDQAKLNVEMGKIHANQYRALIVGTADGINKFTSDATAAGVYKRETDNSG 836
QY 775 NLFTTADDIKGVSTVDMSGYLAVVVPVCAKQDQVRVAADTNOKADGK-SLKTSALDSQ 833
DB 837 VLTFGANDIKGYETFDMSGFVAVVVPVGASNDQIRVAPSTPAKKEGELTLKATEYDSQ 896
QY 834 VIYEGFSNFQD---ANNDADYTNKKIAENADFFKKLGITSPEMAPOYVSATDGSFLDSII 891
DB 897 LIYEGFSNFQITPDGSDPSVVTNRKIAENVDLFKSWGVTSPEMAPQVSADDTGFLDSVI 956
QY 892 QNGYAFSDRYDLAMSKNNKYGSKDDLALNALKALHANGIOATADWPDQIYQLPGEVVTVA 951
DB 957 QNGYAFADRYDLAMSKNNKYGSKEDLRDALKALHKAGIQATADWVPDQIYQLPGEVVTVA 1016
QY 952 KRTNSYGNPTFDAYINNALYATNTKSGSDVQAOYGGAFDLDELKAKYKPMDFVTNNMISTGK 1011
DB 1017 TRTDGAGRKIDAIIDHSLYVANSKSKDQYAKYGGFEFLAEKAKYKPMDFVTNNMISTGK 1076
QY 1012 PIDPSTIKIOMEAKYFNGTINVLGKAGYVLSDDATGKYFTVNNENGDFLPSATGPDQNAKT 1071
DB 1077 PIDDSVKLQWKAKEYFNGTINVLERGVYVLSDEATGKYFTVTKEGNFIPQLQTGKEKVT 1136
QY 1072 GPYDGTGMAYYTSNGKAVNSFIYEGHYHYFYFDKGHWYKSYKABDNDYI-FLPNCI 1130
DB 1137 GFSDDGKITTYFGTSGTOAKSAFVTFNGTYFYFDARGHMTNSEYSPNGKQVYRFLPNCI 1196
QY 1131 QMRDAIYODAGNSYVYGRGTGILYKGDNWPVDPDPNNANK-----TVFRYFDANNVMA 1183
DB 1197 MLSNAFYIDANGNTLYLNSKQMYKGG--YTKFDVSETDKDGKESKVVKFRIFTNEGMA 1254
QY 1184 IGYRNMVQTYFYDENGFAQKQLTDDKTHYFDENGDGAMAKNFVNVGDDWYMDGNG 1243
DB 1255 KGVTVIDGTYFYDENGFAQKQLTDPKGTYYFDHTNGIKDTWRNKGWYFDANG 1314
QY 1244 NAVKQYVNNQILLYFNPETGVQVKGQFITDAQGRTSYDANSKALKSSGFFTPNGSDWY 1303
DB 1315 VAATGAQVINGQKLYFN-EDGSQVKGGVVKNADGTYSKYKGFGBELVTNEFFTTDGNVY 1373
QY 1304 YAEVGYVYKQVQVAENQWYYPQTTGKQAKGAQKVDGRDLYFNPDSGVQVKGDFATD 1363
DB 1374 YA-----GANGKTVTGAQVINGQKLYFNAD-GSQVKGGVKN 1409
QY 1364 ESGNTSFYHGDNGDKVVGSGFFTTGNNAWYADNNGNLVKGFQEIIDGKWHYDDEVTTGQAK 1423
DB 1410 ADGTVSKYNASTGERLTNEFFTTGDNWYIYGANGKSVTGEVKI-----1453
QY 1424 GAALVNGQOLFPDVDSGQVKGDFVTDGQGNITSYDVSNGDKKKVNGFTTGTGNANWYADG 1483
DB 1454 -----GDDTYFFAKDGKQVKGQTVSAGNRIISYTYGDSGKRAVSTWIBIQPVVYVYFDK 1507
QY 1484 QG 1485
DB 1508 NG 1509

RESULT 4
ADX37283
ID ADX37283 standard; protein; 1518 AA.

XX AC ADX37283;
 XX DT 21-APR-2005 (first entry)
 XX DE Streptococcus salivarius glucan binding protein B.
 XX KW immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 XX OS Streptococcus salivarius.
 XX PN US2005031633-A1.
 XX PD 10-FEB-2005.
 XX PF 09-MAR-2004; 2004US-00797821.
 XX PR 13-APR-1998; 98US-0081550P.
 PR 08-JAN-1999; 199US-0115142P.
 PR 12-APR-1999; 99US-00290049.
 PR 07-MAR-2002; 2002US-0363209P.
 PR 08-AUG-2002; 2002US-0402483P.
 PR 07-MAR-2003; 2003US-00383930.
 XX (SMIT/) SMITH D J.
 PA (TAUB/) TAUBMAN M A.
 XX Smith DJ, Taubman MA;
 PI WPI; 2005-151644/16.
 XX
 XX New composition comprising a fragment of a glucan binding protein-B
 PT (GbpB) that binds to MHC class II protein, and a biocompatible
 PT microparticle, useful for producing an antibody (claimed) for immunizing
 PT mammals against dental caries.
 XX Claim 7; SEQ ID NO 40; 73pp; English.
 XX
 XX The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus salivarius GbpB protein of the invention.
 XX Sequence 1518 AA;
 SQ
 Query Match 46.4%; Score 3818; DB 9; Length 1518;
 Best Local Similarity 50.1%; Pred. No. 1.6e-193;
 Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
 QY 1 MEKKLHYKLHKVKKHVTIIVASIGLVSLVG-----AGTVSA---EDKVANDTTAAQTAVG 52
 DB 1 MENKTHYKLHKVKKQWVTIIVASVALATVLGSLVTTSSVSADETQDKTIVTQSNSTTAS 60
 QY 53 VDTGQDQATNTDANTNTDADQSANNTQDQAGSDQSNQDQAKQDQANTDRQADNS 112
 DB 61 LVTSPEATEKADKRNTEKADVLTTPAKETNAVETAT-TNTQATAEATATTADVA-VA 118
 QY 113 QTDNNQATQDQATSPATDGTSGVQRDA---ANVATAADQEGQAPSEQEKSAALSILDNVKL 169
 DB 119 AVPNKEAVVTTDAPATTEKAEQPAVKAEVNVTEVKAPAAALXDSVEAALSILKN 178
 QY 170 IDGKYYVQADGSKYKNFAITVNGQMLYPSDSTGALSSTSTVSFGQTNLVDDPSSHNK 229
 DB 179 IDGKYYVNEIDGSHKENFAITVNGQLLYFGKD-GALTSSTYSFTPGTTNIVDGFSSINR 237
 QY 230 AYDSTAKSFELVNGYLTANSWVRPAGILRNQGTWEASNENDLRPVLMSWTPDKDTQVAVY 289
 DB 238 AYDSSEASFELIDGVLTDADSWVRPASIIKDGVTWQASTAEDRPRLMLMAWPNVDTVQVNYL 297
 QY 290 NYMNYLSANETEVNTNETSQVDLNKEAQSIQTKIEQKITSNDNSTQWLRTAMEAFVAAOPK 349

DB 298 NYMSKVFNL-DAKYSSDTQKQETLKVAAKOIQIKIEQIOAEKSTQWLREITSAFVKTPQ 356
 QY 350 WNMSTENFNKG---DHLQGGALLYTNSDLTPWANSYRLNLRTPPTQDGT--KKYFTTEGG 404
 DB 357 WNKETENYKGGGEDHLQGGALLYVNDSTRTPWANSYRLNLRATATNQGTIDKSLDEQS 416
 QY 405 E----GGYFELLSNDVNSPVVQABQLNQLHYLMWGDIVMGDKDANFGVRVDAVDN 460
 DB 417 DPNHMGGFDELLANDVDLSNPVQAEQLNQIHYLMWGSIVMGDKDANFGIRVDAVDN 476
 QY 461 NADLLQVSNYFKDNYKVTDSANALAHISILSAWSLNDNQYNEEDNGTALSINSSRLT 520
 DB 477 DADMLQLYTNYPREYGYGVNKSANALAHISVLEAWSLNDNHNDKTDGAALAMENKQRLA 536
 QY 521 SLAVLTK-----QPGQRIDLSNLISESVNKERAND-----T 551
 DB 537 LLFSLAKPIKERTPAVSPLYNNTNTTQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNE 596
 QY 552 AYGDPTIPTYSFVRAHDSVQTVIAKIVKESKIDTNSDGYTFTLDOLKDAFKIYNEDMAKVN 611
 DB 597 KYGDASGNTVFIHNDNVQDIIAEIIKKEINPKSDGFTITDAEMKQAEFIYNKMLSSD 656
 QY 612 KTYTHYNIPAAVALLSNMESVPRVYVYDLYTDDGOYMAKSPYDAIATMLQGRYAVS 671
 DB 657 KKYTLNIPAAVAVMLQNMETITRVYVYDLYTDDGHYMETKSPYDITVNLMSKRIKYS 716
 QY 672 GQGSSEVH-----KVGNNQILSSVRYQDILMSADDTQGTDLRSRTSGLVTLV 718
 DB 717 GQGAQRSYWLPDTGKMDNSDELRYTNEVYTSVRYGKIDMTANDTEGSKYRTSGQVTLV 776
 QY 719 SNDPNLDLGGDS-LTVNMGRAHANOAYRPLILGTGKGVSYLKQSD---TNIVKYTDANG 774
 DB 777 ANNPKNLNDQSAKLNVEMKIHANOKYRALIVGTADGIKNFTSDADAIAGYVKETDSNG 836
 QY 775 NLTFTADDIKGYSTVDMSGYLAVWPVGVGAQGDQVRVAADTINQKADGK-SLKTSAALDSQ 833
 DB 837 VLTFGANDIKGYETFTDMSGFVAVWPVGVASDNQDIRVAPSTEAKKEGELTLKATEYDSQ 896
 QY 834 VIYEGFSNPDF--ANNDADYTNKIAENADRFKLGITSPEMAPQVYVATDGSFLDSII 891
 DB 897 LIYEGFSNFQTTIPDGSPPSVYTNRKIAENVDLFKSMGVTSFEMAPQVFSADDTFLDSVI 956
 QY 892 QNGYAFSDRYDLAMSKNNKYGSKDDLANALKALHANGIOAIADWPDQIYQLPGERVFTA 951
 DB 957 QNGYAFADRYDLAMSKNNKYGSKEDRLDALKALHAKGIOAIADWPDQIYQLPGERVFTA 1016
 QY 952 KRTNSYGNPTFDAYINNLYATNTKSSGSDYQAYCGGAFDELKAKYPMFTVNMISTGK 1011
 DB 1017 TRTDGAGRKIADAIIDHSIYVANSKSGKDYQAYGGEFLAELKAKYPMFKVNMISTGK 1076
 QY 1012 PIDPSTKIQWEAKYFNGTNLGKAGYVLSDDATGKYFTVNVENGDFLPASFTGDQNAKT 1071
 DB 1077 PIDDSVKLKQWKAEBYFNGTNLVLERGVYVLSDEATGKYFTVTKEGNFIPLQLTGKEKVI 1136
 QY 1072 GPYVDTGTMAYYSTSGNKAVNSFIYEGGHYVYFDKGHMTGYSYKAEQDNDYY-FLPNCI 1130
 DB 1137 GFSSDGKGIYFTGTSQAKSAFVTFNGNTYTFDARGHMTVNSEYSPNGKDYVRFPLNGI 1196
 QY 1131 QMRDAIYQADQNSYYYIGRTGILYKGDNMWPPVDPNNANK-----TVFRYFDANNVMA 1183
 DB 1197 MLSNAFYIDANGNTLYNSKGMQYKGG--YTKFDVSETDKDGESKVKYKFRFTNEGVA 1254
 QY 1184 IGRNMVQTYTFDBNGFOAKQLTDDKTHYFDEDNAMAKNFVNVDDWYMDGNG 1243
 DB 1255 KGVTVIDGFTQYFGBDGFQAKDKLVTFKGTIYFFDAHTGNGIKDTRNKGWYFDANG 1314
 QY 1244 NAVKQYVNNQILYFNEBETGVQVKGQFITDAQGRTSYYDANSGLKSSGFFTPNGSDWY 1303
 DB 1315 VAATGAQVINGOKLYFN-EDGSQVKGVVKNADGTYSKYKEGFLVTFNEFTTGNVWY 1373
 QY 1304 YAEANGYVYKFKQVAENQDQWYFYDQTTGKQAKGAQKDRDLYFNPSGQVGVKGDPAFD 1363

Db 1374 YA-----GANGKTYTGAQVINGOHLFYFAD--GSQVKGGVK 1409
Qy 1364 ESGNTSFYHGDNGDKVVGFFFTGNNAYADNNGNLVKGFOEIDGKWHFDEV7GQQA 1423
Db 1410 ADGTYSKYNASTGERLTNEFFTTGDNWYYIGANGKSVTGEVKI-----1453

Qy 1424 GAALVNGQQLYFDVDSGIOVKGFYTDGOGNTSYDVNSGDKKXNGFFTTGDNWYYADG 1483
Db 1454 -----GDDTYFAXDKGQVKQTVSAGNGRISYYGDSGKRAVSTWIEIQPGVYVFDK 1507

Qy 1484 QG 1485

Db 1508 NG 1509

RESULT 5

AAU98029

ID AAU98029 standard; protein; 1430 AA.

XX AAU98029;

XX

DT 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFD.

XX Glucosyltransferase; GTFD; transgenic plant; paper sizing;

XX coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture.

XX Streptococcus mutans.

XX US2002031826-A1.

PN 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX N-PSDB; ABK52940.

XX Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.

XX Disclosure; Page 38-42; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from I448V, D457N, D567T,

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,

CC I448V/D457N/D567T/D571K/K799Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked to a promoter, a vector comprising the expression cassette, host

CC cell introduced with the vector, a transgenic plant comprising the

CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or

CC coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GTF, wild type or, starch, a latex,

CC thermoplastic molecule or their combinations or glucan and starch where

CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents GTFD

XX SQ Sequence 1430 AA;

Query Match 43.9%; Score 3615; DB 5; Length 1430;

Best Local Similarity 49.9%; Pred. No. 9e-183;

Matches 745; Conservative 217; Mismatches 418; Indels 112; Gaps 32;

Qy 1 MEKHLHYLKHVKGHVTVIAVASIGLSLVGAGTVSAEDKVDVDTAQTAVGVDTGQ---57

Db 1 METKERYKMHVKGHVTVIVAS--GLITL---GTTTLGSSVSAETEQQTSKVVTKQSD 56

Qy 58 DOATTNDANTVTTDTADQSANNTNQDQSGDSNNQQAQKDT-----ANTDR 106

Db 57 DKAASESSQTDPKPT---KQAQTEQTAQS--QANVADTSTSTIKETPSQNTTQANSDD 111

Qy 107 NQADNSQTDNQAQDQATSPATDGTSVQRDAANVATA--ADQEGQTAPSEQSKSAAL--SL 164

Db 112 KTVNTKSEEAQTSERTKQSEEAQTASSQALTOAKAELTKQRTAQAKNKPVDLAAI 171

Qy 165 DNVKLIDGKYVYQADGSKYKNFAITVNGQMLYFDSDTGALSSTSYSGFQQTTLVDVDF 224

Db 172 PNVKQIDGKYVYIGSDGQPKKFNALTNNKLVLYFDKNTGALTDTSOYQFKQGLTKLNNDY 231

Qy 235 SSHNKAYDSTAKSFELVNGYLTANSWYPPAGILRNGQTWEASNDLIRPLMSWHPDKDT 284

Db 232 TPNHQIVNFENTSLETIDNYVTADSWYRPPKQILKNGKWTWTASSESDLRPLMSWHPDKDT 291

Qy 285 OVAVYVYNNKYLSANETETVNTNETSQVDLNKEAQSIQTKEQKITSDNSQWMLRTAMEAFV 344

Db 292 QIAYLVNYYNQGLGTGENYTADSSQESLNLAQTVQVKIETKISQTOQTOWLRDIINSFV 351

Qy 345 AAQPKNNMSTE---NFKNGDHLQGGALLYTNSDLTPWANSDYRLNRRTPQDGTGKYFT 401

Db 352 KTQPNWNSQTESDTSAGEKDHQGGALLYSNSDKTAYANSDYRLNRRTPSTQGTGPKYPE 411

Qy 402 EGGEGGYEFLSNVDNSNPVQARQINQLHVLNMGDIYVMGDKDANDEGVAVDANNV 461

Db 412 DNSSGGYDFLLANDIDNSNPVQAEQLNWLHYLMNYGSIIVANDPEANFDEGVAVDANNV 471

Qy 462 ADLLQVYNNYKDYKVTVDSEANLAHISILEAWSLNDNQVNETNGTALSIDNSRLTS 521

Db 472 ADLLQIAGDYLAHGVGVKSEKNAIHLISLEAWSNDPQYNKDTKGAGLPIDNKLRLSL 531

Qy 522 LAVLTKQPCQRIDLNS-----LISESVNKERANDTAYGDTIPTYSFVRAHDSVQ 571

Db 532 LVALTR-PLEK-DASNKNEIRSGLEPVITNSLN-NRSAEGKNSRMANYIFIRAHDSVQ 588

Qy 572 TVIAKIVKEIDTNSDGTFTFLDQIKDAFKYINEDMAKVNKYTYHYNIPAAALLISNME 631

Db 589 TVIAKIIHQAQINPKPTDGLTFTLDELQAPKIYNEDMRQAKKYTOSNIPTALMLSNKD 648

Qy 632 SVPRVYVGDLYTDDCGYMAKSPYYDAITATMLQGRYAVVSGGQSEVHKVGNQ-----686

Db 532 LVALTR-PLEK-DASNKNEIRSGLEPVITNSLN-NRSAEGKSKERMANVIFPRAHDSEVQ 588
 QY 572 TVIAKIVKEKIDTNSDGYFTLQDLKDAFKINEMAKVNKTYTHYNIPAAVALLSNME 631
 Db 589 TVIAKIIKAQINPKTDGLTFTLDELKQAFKINEDMRQAKKKTQSNITAYALMLSNKD 648
 QY 632 SVPRVYGYDLYTDDGOYMAKKSPYDAIATMLQGRITAYVSGOSEVHKVGNQ----- 686
 Db 649 SITRLYGYDMYSDDGQYMATKSPYDAITLLKARIKYAAGQDMKITTYVEGDKSHMDWD 708
 QY 687 ---ILSSVRYGQDLMSADTQGTDLRSRTSLVTLVSNPNLDLG-GDSLTVNMGRAHAKQ 742
 Db 709 YTGVLTSVRYGYGANEAED-QQSEATKTQGMAVITSNPSLKLNQNDKVLVNMGAHAKQ 767
 QY 743 AYRPLILGTGQVQSYLKSDT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPV 801
 Db 768 EYRPLLLTTKDLGTSYSDAAKSLYRKTNDRKDELVDASDIQGYLNPQVSGYLAVWVPV 827
 QY 802 GAKGDQVRAADTNQKADGKSLKTSAAALDSQVIVYEGFSNPODPANNADYTNKKIAENA 861
 Db 828 GASDNQDVRVAASNKANATQVYESSALDSQLIVYEGFSNPODPVTKDSYTNKKIAQNV 887
 QY 862 DPFKLGITSFEMAPQVYSATDGSFLDSIIQNGYAFSDRYDLAMSKNKKYGSQDDLANAL 921
 Db 888 QLFKSGVTSFEMAPQVYSSEDSFLDSIIQNGYAFEDRYDLAMSKNKKYGSQDDMINAV 947
 QY 922 KALHANGIOAIADWPDIQYQLPGEVVTAKRTNSYGNFTFDAYINNALYATNTKSSGD 981
 Db 948 KALHKSIGIQUIADWPDIQYQLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTKNGKD 1007
 QY 982 YQAOYGGAFDELKAKYPMFTVNMIISTGKPTDPTKIKQWEAKYFNGTNIIVLGAGYVL 1041
 Db 1008 YQAOYGGAFSELAAKYPIFNRTQISNGKKIDPSEKITAMKAKYFNGTNIIVLGRGVYVL 1067
 QY 1042 SDDATKYFTVNEGDFLPASFTGQONAKTGFYDGTGMAYTSTSNKAVNSFIYEG-GH 1100
 Db 1068 KDNASKYFELKGNQTYLPKQWT-NKEASTGFVNDGNGMTFTSTGYQAKNSFVQDAKN 1126
 QY 1101 YYYFDKDHMTVGSYKAEADGNDYFPLNGIQMRDAIYDQAQNSYNYTGTILYKGDWNY 1160
 Db 1127 WYFFDNNGHMYGLQQL-NGEVOYFLSNGVQLRESFLENADGSKNYFGLGNRY-SNGYY 1184
 QY 1161 PFVDPNNAKTVFYPDANNVAIGYRMYGQTYFDFDENGFOAKGOLLTDDKG-THYFDE 1219
 Db 1185 SF-----DNDKRWYFDASGVMAVLKTTINGNTQVFDQDGYQVKGAWITGSDGKRYFDD 1239
 QY 1220 DNGMAKNKPVN-VGDDWYMDGNGNAVKGQYVNNQILYFNPETGVQVKGQFITDAQGR 1278
 Db 1240 GSGNMAVNRFANDKNGDWYLSNDSGIALVGVQTINGKTYFG-QDGKQIKGLIITD-NGK 1297
 QY 1279 TSYDANSALKSSGFTFPGNSDWYIAENGYYVYKFKQVAENQDOWYYPDQTGKQAKGA 1338
 Db 1298 LKYFLANSGLARNIFAT-----DSQNNWYFYG-SDGVAVTGS 1334
 QY 1339 AKVGDGLYFNPDSGVQVKGDFATDESQNTSFPHGNDGKVGPGFTTGNNAWYADNNG 1398
 Db 1335 QTIAGKGLYFASD-GRQVKGFSFT-YNGKVHYTHADSGELQVNRFPADKGNWYLDNSG 1392
 QY 1399 NLVKGQEIIDGKWHYHDFEVTGQAKGAALVNGQLYFVDVSGIYQVKGDFVTD 1450
 Db 1393 EALTGSQRI-----NDQRVFF-TREGKQVKGVDAYD 1422

RESULT 7

ADX37279

ID ADX37279 standard; protein; 1430 AA.

XX AC ADX37279;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus mutant glucan binding protein B variant #8.

XX immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 XX Streptococcus mutans.
 XX US2005031633-A1.
 XX 10-FEB-2005.
 XX 09-MAR-2004; 2004US-00797821.
 XX 13-APR-1998; 98US-0081550P.
 XX 08-JAN-1999; 99US-0115142P.
 XX 12-APR-1999; 99US-00290049.
 XX 07-MAR-2002; 2002US-0363209P.
 XX 08-AUG-2002; 2002US-0402483P.
 XX 07-MAR-2003; 2003US-00383930.
 XX (SMIT/) SMITH D J.
 XX (TAUB/) TAUBMAN M A.
 XX Smith DJ, Taubman MA;
 XX WPI; 2005-151644/16.
 XX New composition comprising a fragment of a glucan binding protein-B
 XX (GbpB) that binds to MHC class II protein, and a biocompatible
 XX microparticle, useful for producing an antibody (claimed) for immunizing
 XX mammals against dental caries.
 XX Claim 7; SEQ ID NO 36; 73pp; English.
 XX The invention relates to a composition comprising a fragment of a glucan
 XX binding protein-B (GbpB) and a biocompatible microparticle, where the
 XX fragment binds to a major histocompatibility complex (MHC) class II
 XX protein. The composition is useful for producing an antibody for
 XX immunizing mammals against dental caries. This sequence corresponds to a
 XX Streptococcus mutans GbpB protein of the invention.
 XX Sequence 1430 AA;
 Query Match 43.9%; Score 3615; DB 9; Length 1430;
 Best Local Similarity 49.9%; Pred. No. 9e-103;
 Matches 745; Conservative 217; Mismatches 418; Indels 112; Gaps 32;
 QY 1 MEKHLHYKLHKYKHWVTIAVASIGLVSIVGAGTVSAEDKVANDTQAATVGVDTGQ---- 57
 Db 1 METKERYKMKYKHWVTIAVAS-GLITL---GTTTLGSSVSAETEQQTSKVVTKSBD 56
 QY 58 DOATTNDANTNTTDDTADQSANNTNQDQAGSDQSNQDQAKQDT-----ANTDR 106
 Db 57 DKAAESSQTDAPKT---KQAQTEQTAQS-QANVADTSTSTITKTPSQNTTQANSDD 111
 QY 107 NQADNSQTDNQATQATSPATDQSVQRRDAANVATA-ADQEGQTAPSEQESKAAL-SL 164
 Db 112 KTVNTKSEEAQTSBERTKQSEEAQTASSQALTKAKELTKQRTAAQENKPNVDLAAI 171
 QY 165 DNVKLIDGKYVYQADGSGYKKNFALTIVNGQMLYFSDTGALSSSTYSFSQGTNLLVDDF 224
 Db 172 PNVKQIDGKYVYIGSDGQPKKNFALTIVNKVLYFDKNTGALTDTTSQYQFKQGLTKLNDY 231
 QY 225 SHNKAYDSTAKSFELVNGYLTANSWYRPAGILRNGQTVWEASNDLRLPVLMSWPKDT 284
 Db 232 TPNQIVNPENTSLTIDNYVTADSWYRPKDLKNGKTWTASSESDRLPFLMSWPKDT 291
 QY 285 QVAYVYNNKYL SANETETVNTSQTDLNKEAQSQTKEOKITSDNSTQWLRLTAMEAFV 344
 Db 292 QIAYLNTNQQLGTGENVTADSSQESLNLAATQVQVLETKISQTOQTQWLRLDINSFV 351
 QY 345 AAQPKWNMSTE---NFKRGDHLQGGALLYNSDLTPWANSDYRLNRLTPTQODGTGKYFT 401
 Db 352 KTQPNWNSQTESDTSAGEKDHQGGALLYNSDKTAYANSDYRLNRLTPTSGPKPYE 411

QY 462 ADLLQVSYNFYKDYKVTSEANALAHISILEAWSLNDNQYNEEDTNGTALSIDNSRLTS 521
 Db 472 ADLLQVSYNFYKDYKVTSEANALAHISILEAWSLNDNQYNEEDTNGTALSIDNSRLTS 531
 QY 522 LAVLTQPCQRIDLSN-----LISESVNKERANDTAYGDTTPTYSFVRAHSEVQ 571
 Db 532 LVALTR-PLER-DASNEIRSGLEPVITNSLN-NRSABGKNSERNWYIFIRAHSEVQ 588
 QY 572 TVIAKIVKEKIDTNSDGYFTTLDQLDAFKIYNEEDMAKVNKYTHYNIIPAAVALLSNME 631
 Db 589 DVIAKIINKAIPKTDGLTGLTDELAKQAFKIYNEEDMRQAKKYQTSNIPYAVALLSNKD 648
 QY 632 SVPRVYDGLYTDGQYMAKSPYDAIATMLGRIATYVSGGSEEVHKVNGNQ-----686
 Db 649 SITRLYDGMYSDDGYMATKSPYDAIDTLKARIKAAAGQDMKITTYVEGDKSHMDW 708
 QY 687 ---ILSSVRYGQDLMSADDTQGLTSLRTSLVTLVSNDPNLDLG-CDLSLTVMNGRAHANO 742
 Db 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPFLKLNQNDKVINVMGAHKNQ 767
 QY 743 AYRPLTLGTGQVQSYLKDSDT-NIVKYTDANGNLTFTADDIKGYSTVDMSGYLAVWVPV 801
 Db 768 EYRPLTLTKDGLTSYTSAAKSLYKTDNGELVFDASDIQGYLNPQVSGYLAVWVPV 827
 QY 802 GAKGDQVRVAADTKQADKSLKTSALDSQVIVYEGFSNFQDFANNDADYTNKKIAENA 861
 Db 828 GASDNQDVRVAASNKANATQGVYESSALDSQVIVYEGFSNFQDFVTKDSYTNKKIAQV 887
 QY 862 DFEKGLGTSPEWAPQVYVATGSGFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLALNAL 921
 Db 888 QLFKSGVTSFENAPQVYVSSGFLDSIIQNGYAFSDRYDLAMSKNNKYGSKQDQMINAV 947
 QY 922 KALHANGIQAIADWPQDQIYQLPGEVVTAKTNSYGNPTFDAYINNALYATNTKSSGSD 981
 Db 948 KALHKSQIQAIDWPQDQIYQLPGEVVTAKTNSYGNPTFDAYINNALYATNTKSSGSD 1007
 QY 982 YQAYGGAFDELKAKYPMFTVMNIMSTGKPIDPSTKIKOWEAKYFNGTNVLGKAGYVL 1041
 Db 1008 YQAYGGAFDELKAKYPMFTVMNIMSTGKPIDPSTKIKOWEAKYFNGTNVLGKAGYVL 1067
 QY 1042 SDAATGKYFTVNEGDFLPASFTGQONAKTGYDGTGMAYYSTSGNAKAVSFYEG-GH 1100
 Db 1068 KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFTYSTGYQAKNSFVQDAKEN 1126
 QY 1101 YYYFDKGHMTGVSYKAEDGNDYFLPNGIQMRDAIYQDAQNSYVYGRITGLYKGDNMY 1160
 Db 1127 WYFPDNGHMYGLQOL-NGEVOYFLSNGVQVRESFLENADGSKNYFGLGNRY-SNGY 1184
 QY 1161 PFVDPNNANKTVRYFDANNVMAIGYRNMYGQTYTFDENGFOAKGQLLTDKRG-THYFDE 1219
 Db 1185 SF-----DNDSKWRVYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSGKKRYFDD 1239
 QY 1220 DNGAMAKKFFVY-VGDDWYMDGNGNAKGVQPVNNQILYFNPETGVQVKGQFITTDAQR 1278
 Db 1240 GSGNMAVRFANDKNGDMYTLNSDGIALVGQVQTINGKTYFYG-QDGKQIKGKIITD-NGK 1297
 QY 1279 TSYVDANSALKSSGFFTPNGSDWYAEYVYKGFQVAENQDQWYFDQTTGQAKGA 1338
 Db 1298 LKYLFLANGELARNIFAT-----SDQNWYVFG-SDGVAVTGS 1334
 QY 1339 AKVGDRLYFNPDSGVQVKGDFPATDESNGTSFYHGDNGDKVVGFFFTGNNAWYADNNG 1398
 Db 1335 QTIACKKLYFASD-CKQVKGSPVT-YNGKVHYHADSGELQVNRFEADKDGWYLLDSNG 1392
 QY 1399 NLVKGQFQIDGKWHYFDEVTGQQAAGALVNGCQQLYFVDVDSGIQVKGDFVTD 1450
 Db 1393 EALTGSQRI-----NDQRVFF-TREGQVKGQDVAYD 1422

RESULT 11
 AAU98044
 ID AAU98044 standard; protein; 1430 AA.
 XX

AC AAU98044;
 XX 27-AUG-2002 (first entry)
 DT
 XX S. mutans glucosyltransferase GTFD mutant N471D/T589D.
 XX Glucosyltransferase; GTFD; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX Streptococcus mutans.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 471 /note= "wild-type Asn substituted by Asp"
 FT Misc-difference 589 /note= "wild-type Thr substituted by Asp"
 FT
 XX US2002031826-A1.
 PN
 XX 14-MAR-2002.
 PD
 XX 19-DEC-2000; 2000US-00740274.
 PF
 XX 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX (NICH/) NICHOLS S E.
 PA
 XX Nichols SE;
 PI
 XX WPI; 2002-414332/44.
 DR
 XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 PS Claim 36; Page; 44pp; English.
 XX
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 1448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes PI or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch


```

||||:  :: ||| ||| : : ||| ||| ||| : : ||| ||| : : ||| :
698 TTISGVGVGQNPFSLDGKVVALNMGAAHQEYRALMVKTKGQVATYADASKAG 757
QY
766 IVKYTDANGLTFTADDIKGYSTVDSMGYLAVMVPVGAKGDQDVRVAADTNQKADKSLK 825
Db
758 LVRKTUENGYYLFLNDLKGAVNPQVSGFLQVWPVGAADDQDIRVAASDTASTDGKSLH 817
QY
826 TSAALDSQVIYEGFSNFPQANNADYTNKKTAENADPFKKLGITTSFENAPQVVSATDGS 885
Db
818 QDAAMDSTRVMFEGFSNFPQSFPATKEEYTNVAVANNVDFVSWGIIITDFEMAPQVVSSTGQ 877
QY
886 FLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLALAKHALHANGIOAIAADVPDQIYOLPG 945
Db
878 FLDSVLQNGYAFIDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMTFFPK 937
QY
946 EEVTAKTNSYGNPTFDYAINNALYATNTKSSGSDYQAYGGAFDELKAKYPMFTVN 1005
Db
938 QEVVTVTRTDKFGKPIAGSQINHSLVYTDTKSSGDDYQAKYGGAFDELKAKYPELFTKK 997
QY
1006 MISTGKPIDPSTKIKQWEAKYFNGTNVLKGAGYVLSDDATGKYFTVWNGDFLPASTFG 1065
Db
998 QISTGAIDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVASDTLFLPSLLG 1057
QY
1066 DQNAKTGFYDGTGMAYYST-SGNKAVNSFIYEGGHYYFDKDHMTVGSYKAEDGNDY 1124
Db
1058 -KVESGIRYDGRGYINSSATGQDVKASFIIEAGNLYYFGKGYMTGA-QTINGANFY 1115
QY
1125 FLPNGTQMRDALYQDAQNSYYVYRTGILYK-----GDMYFPVDPNNAKNTVRYFDA 1178
Db
1116 FLENGTALANTYTDAGNSHYIYANDGKEYNGYQFGNDW-----RYFKD 1161
QY
1179 NNVMALGYRNMGTQYTFYFENGQAKQLL-TDDKGTHTYFDESDNGMAKNKV-NVGGDW 1236
Db
1162 GN-MAVGLTAVDGNVQYFDKDGQVQAKKIIVTRDGRVRYFDQHNGNAATNTFIADKTGH 1220
QY
1237 YVMDGNAGNAVGOYPVNNQILYFNPETGVQVKGQFITDAQRTSYVDVDSNGALKSSGFFT 1296
Db
1221 YYLKDGAVATGAQTVGKQLYFE-ANGQQVKGDFVTSDEGLYFYDVSDDMTDTFIE 1279
QY
1297 PNGSDWYIYAEYVYKGFQKQVAENQDQWYTFDQTTGKQAKGAKVDGRDLYFNPDSGVQV 1356
Db
1280 DRAGNWFYLGK-----DGAAVTGAQTIRGQKLYFKA-NGQQV 1315
QY
1357 KGFADDESNTSYFHYGNDGDKVGGFFTTGNNAWYIADNG-----NLVKG----- 1403
Db
1316 KGDIVKGTGDKIRYDAKSGEQVFNKTVKADGKTYVIIGNDGVAVDPSVYKGTQFKDASG 1375
QY
1404 ---FQIDCK-----WY---HFDEV---TGQQAAGALVNGQQLYFDVDSGIVQKGFV 1448
Db
1376 ALRFYNLKQLVTGSGWYETANHDWVYIQSGKALTGEQTINGQHLFYKED-GHQVKGQLV 1434
QY
1449 TDCQNTSYDVSNGDKVNGFFTTGDNAWY----- 1480
Db
1435 TGTGDKVRYIYDANGSDQAFNKSVTYNGKTYIFGNDGTAQTAGNPKGQTFKDGSDIRFYSM 1494
QY
1481 -----ADGQ-----GNLAKGRKSIDNQDLTFDPATGKQVKGQLV-SIDGRN 1520
Db
1495 EGQLVTGSGWYENAOQWLYVXNGKVLTLGLTVGSRQVYFD-ENGIOAKGKAVRTSDGKI 1553
QY
1521 YVFDSSGNNMKNRFVIRIGDQWYIFGNDGA 1550
Db
1554 RYFDENSGSNITNQWKFVYQYIYFNDGA 1583

```

RESULT 15

AAU79284

ID AAU79284 standard; protein; 1476 AA.

XX AAU79284;

AC

XX 13-AUG-2002 (first entry)

XX Streptococcus mutans monoclonal antibody-related protein #1.

DE

XX Antibody; dental caries; water insoluble glucan synthetase; anti-caries;
 KW glucosyl transferase-B; immunotherapy.
 XX Streptococcus mutans.
 OS JP2002114709-A.
 PN 16-APR-2002.
 PD 04-OCT-2000; 2000JP-00304889.
 PF 04-OCT-2000; 2000JP-00304889.
 PR 04-OCT-2000; 2000JP-00304889.
 XX (UYN1-) UNIV NIPPON.
 PA WPI; 2002-448885/48.
 DR XX
 XX Anti-caries agent composed of a monoclonal antibody against an inhibitory
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-
 PT B (GTF-B) of Streptococcus mutans.
 PS Claim 3; Page 13-16; 28pp; Japanese.
 XX The invention relates to a monoclonal antibody against dental caries and
 CC an anti-caries agent composed of a monoclonal antibody produced by 17566)
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl
 CC transferase-B. The monoclonal antibody specifically inhibits water
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
 CC transferase-B and is used in the immunotherapy of dental caries. This
 CC sequence represents a Streptococcus mutans monoclonal antibody-related
 CC protein
 XX Sequence 1476 AA;

Query Match 42.5%; Score 3497; DB 5; Length 1476;
 Best Local Similarity 47.1%; Pred. No. 1.7e-176;
 Matches 737; Conservative 246; Mismatches 457; Indels 126; Gaps 37;

QY 1 MEKLLHYKLHKYKHWVTIAVAS-IGLVSLVGAGTVAEDKVANDTTAQTATVGVDTGQDQ 59
 Db 1 MDKVKYKLRKVKKRWTVSVASAVMTLTLSGGLVKAD---SNEKSQ----- 46
 QY 60 ATTDANTNTTDTDTADQSAN-----TNQDQAGSQNNQDQAKQDTAN----- 103
 Db 47 -ISNDSNTSVV---TANESNVITEATSKQEAASSQTNHTVTTSSTSVVNPKEVSNP 102
 QY 104 -TDRNQADNSQTDNNQATDQATSPATDGTSVORRDAANVATAADQEGQTAPSEQEKSAAL 162
 Db 103 YTVGETASNGEKLQNTT-----TVDKTS-----EAAANNISKQTEADTDVDDGSAANL 153
 QY 163 ----SLDNNVCLIDGKYYVQADGSKYKNFAITVNGOMLYFDSGDTGALSSTSYVSQGT-T 217
 Db 154 QLEKLPVKEIDGKYYDNNKGVRTNFTLIADGKILHFD-ETGAYTDTSDITVKNQDIV 212
 QY 218 TNLVDDFSSHNAKYDSTAKSFELVNGYLTANSWYRPAQILNRGTQWTEASNENDLRPLVMS 277
 Db 213 TTRSPLYKKYQVYDRSAQSFPHVHYLTAEWSYRPFYILDKGTWTQSTEKDPRLLMT 272
 QY 278 WMPDKDTQVAYVYNNKYLSANETEVNETSQVDLNKEAQSIQTKEQKITSDNSTQWL 337
 Db 273 WMPDQETQRYVYNNYNAQLGINKT-YDDTSNQLNLNIAAATIAQAKIEAKITTLKNTDWLR 331
 QY 338 TAMEAFVAAQPKWNMSTEN-FNKGDHLOGGALLYTN-SDLTPWANSDYRLNARTPTQDQ 395
 Db 332 QTISAFVKTQSAWNSDSEKFPD--DHLQNGAVLYDNEGKLTTPYANSYRILNRTPTNQ 389
 QY 396 TK--KYFTGGEGGEGYFLLSDNDVNSNPVQAEQNLQHLNMGWDIVMGDKDANPDGVR 453
 Db 390 KCDPRYTADNTIGGYEFLFLANDVDNSNPVQAEQNLNWLHFLMNFNGIYVNDPDANFDSIR 449

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 19:39:30 ; Search time 34.0237 Seconds
(without alignments)
3776.130 Million cell updates/sec
Title: US-10-797-821-38
Perfect score: 8237
Sequence: 1 MEKLUHYLHKYKHVVIA.....FVRIGDQWYFGNGAATNL 1554
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/iaa/5 COMB.pep.*
2: /cgn2_6/prodata/1/iaa/6 COMB.pep.*
3: /cgn2_6/prodata/1/iaa/H COMB.pep.*
4: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
5: /cgn2_6/prodata/1/iaa/RE COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3615	43.9	1430	2	US-09-008-172-2
2	3615	43.9	1430	2	US-09-210-361-6
3	3615	43.9	1430	2	US-09-740-274-6
4	3434.5	41.7	1475	2	US-09-007-999-2
5	3434.5	41.7	1475	2	US-09-210-361-2
6	3434.5	41.7	1475	2	US-09-740-274-2
7	3429.5	41.6	1577	1	US-08-793-824-2
8	3161.5	38.4	1375	2	US-09-210-361-4
9	3161.5	38.4	1375	2	US-09-740-274-4
10	2422	29.4	2057	2	US-09-499-203-2
11	2285	27.7	1278	2	US-09-604-957-3
12	2285	27.7	1278	2	US-09-995-749A-2
13	1643	19.9	545	2	US-09-604-957-4
14	1615	19.6	545	2	US-09-995-749A-10
15	1442.5	17.5	522	2	US-09-995-749A-11
16	1440	17.5	523	2	US-09-604-957-5
17	1365	15.4	535	2	US-09-604-957-7
18	1365	15.4	535	2	US-09-995-749A-13
19	1174.5	14.3	584	2	US-09-604-957-6
20	1172.5	14.2	584	2	US-09-995-749A-12
21	563	6.8	349	2	US-09-009-620-2
22	545.5	6.6	2710	1	US-08-480-604A-6
23	545.5	6.6	2710	1	US-08-405-496A-6
24	545.5	6.6	2710	2	US-08-915-136-6
25	545.5	6.6	2710	2	US-08-957-310-6
26	545.5	6.6	2710	2	US-10-011-366-6
27	545.5	6.6	2710	2	US-09-084-517-6

RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match 43.9%; Score 3615; DB 2; Length 1430;
Best Local Similarity 49.9%; Pred. No. 9.7e-211;
Matches 745; Conservative 217; Mismatches 418; Indels 112; Gaps 32;

QY	1	MEKLUHYLHKYKHVVIAVASIGLVAGAGTVSAEDKVANDTTAQATVGVDTCQ---	57
Db	1	METKERYKHVKYKHVVIAVAS-GLITL--GTTTLGSSVSAETEQTSDKVVTKQSD	56
QY	58	DOATNDANTNTTDTTADQASANTNODQGSQSNNOQAKQDT-----ANTDR	106
Db	57	DKAASESSQDAPKT---KQAQTEQTAQS-QANVADTSTITKETPSQNTTQANSDD	111
QY	107	NOADNSQTDNNQATQATSPADGTSVQRDAANVATA-ADQEGOTAPSEQKSAAL-SL	164
Db	112	KTVNTKSEEAQTSEERTKQSEAAQTASSQALTOAKAELTKQRTAAQENKNFVDLAAI	171
QY	165	DNVKLIDGKYVYVQADGSKYKKNFALTAVNGQMLYFSDTGALSTSTYSFSQGTNVLVDDF	224
Db	172	PNVKIDGKYVYIGSDGQPKKNFALTAVNNKLVYFDKNTGALTDTSQYQPKQGLTKLNDY	231
QY	225	SHNKAYDSTAKSFELVNGYLTANSWYRPAQILRNGQTWASNEENDLRPLVLSWVPDKDT	284
Db	232	TPHNQIVNFENTSLTIDNYVTADSWYRPAQILRNGQTWASNEENDLRPLVLSWVPDKDT	291
QY	285	QVAYVYNNKYLSANETETVNTSVDNKEAQSTQTKIEQKITSDNSTQWLTAMEAFV	344
Db	292	QIAYLYNNQOGLGTGENYTDSSQESLNLAATQVQVQKIETKISQQTQOWLRDIINSFV	351
QY	345	AAQPKWNMSTE---NFNKGDLQGGALLYTNSDLTPWANSDYRLNRTPTQDGTGKTYT	401

ALIGNMENTS

QY	522	LAVLTQPCQORIDLSN-----LISESVNKERANDTAYGDTPIPTYSFVRAHDSVQ	571
Db	532	LVALTR-PLEK-DASNKNEIRSGLEPVTJNSLN-NRSAEGKNSERMANYIFIRAHDSVQ	588
QY	572	TVIAIKVKEIDTNSDGYTFTLDQLKDAFKIYNEDMAKVNTYTHYNIIPAAVALLLSNME	631
Db	589	TVIAIKIQAQINPKTDGLTFTLDELKQAFKIYNEDMROAKKYIQTOSNIPTAYALMLSNKD	648
QY	632	SVPRVYVYGDLYTDDGOYMAKSPYYDAIATMLQGLRIAYVSGQSEBHVHVGNNQ-----	686
Db	649	SITRLYYGDMYSDDGOYMATKSPYYDAITDLLKARIKYAAGQDMKITYVEYGDKSHMDWD	708
QY	687	---ILGSVRYGQDLMASADDTQGTDLSPRTSGLVTLVSNDPNLDLG-GDSLTVNMGRAHANQ	742
Db	709	YTGVLTSVRYGTGANEATD-QGSEATKTQGMAVITSNPSSLKLNQNDKVIYNMGAHKVQ	767
QY	743	AYRPLILGTQDGVQSYLXKDSDT-NIVKYITDANGNLTFETADDIKGYSTVDMSGYLAVMPV	801
Db	768	EYRPLLLTTKGLTSYTSDDAAKSYIRKINDKGLVFDASDIQGLVNPQVSGYLAVMPV	827
QY	802	GAQGDQDVRVAADTNQKADGSKLKTSAALDSQVIYEGFSNFQDFANNDADYTNKKIAENA	861
Db	828	GASDNQDVRVAASNAKANATGVYESSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQNV	887
QY	862	DFFKLGTITSEMAPQYVSATDGSFLDSIIQNGYAFPSDRYDLAMSKNKYSGKDDLANAL	921
Db	888	QLFKSWGVTSEMAPQYVSSYSDGSFLDSIIQNGYAFEDRYDLAMSKNKYSGQDMINAV	947
QY	922	KALHANGIOATADWVPDQIYQIPEGEEVTVAKRTSYGNPTPDAYINNALYATNTKSGSD	981
Db	948	KALHKSQIQTADWVPDQIYNLPKGEVVVATRVNDYGEYRKDSIEIKNTLYAANTKSGKD	1007
QY	982	YQAOYGGAFBLDELKAKYPDMFTVNMI STGKPIDPSTKIKOWEAKYENGTVNLHGKAGVVL	1041
Db	1008	YQAKYGGAFLELAAKYPSIFNRTQISNGKKIDPSEKITAKWAKYFNGTNLHGRGVYVL	1067
QY	1042	SDDATGYFTYVNGENDFLPASFTGDNQAKTFYVYDGTGMAYYSTSGNKAVNSFYEG-GH	1100
Db	1068	KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTPYSTSGYQAKNSFVQDAKGN	1126
QY	1101	YYYFDKQGHMVTGYSKABEDGNDYLYPLNGIOMRDALYYQDAQNSYYYGRTGILYKGDWY	1160
Db	1127	WYYFDNNGHMYGLQQL-NGEVQYFLSNGVLQRESFLENADGSKNYFGLGNRY-SNGYY	1184
QY	1161	PFVDPNNAKTVFYFYFDANNWALGYRNWYQTYFYFDENGPOAKQLLTDQKG-THYFDE	1219
Db	1185	SF-----DNDSKMYFYDASGVMAVGLKTINGTQYFDQDGYQVKGAWITGSDGKKRYFDD	1239
QY	1220	DNGAMAKNKFVN-VGDDWYVMDGNGNAVKGQYPVNNQILYFNPETGVQVKGFOITTDQQR	1278
Db	1240	GSQNMVNRFANDKNGDWYLYNSOGIALUVQVTTINGKTYFYFQ-ODGKQIQKGIITD-NGK	1297
QY	1279	TSYYDANSALKSSGFFTPNGSDWYVYKGFQKVAENQDWYFYFQTTGKQAKGA	1338
Db	1298	LKYFLANSGLARNIFAT-----DSQNNWYFYFQ-SDGVATGS	1334
QY	1339	AKVDGRDLYNPDSGVQKGDFAFDSEGNSTSFYHGDNGDKVVGQFFFTTGNMAYADNNG	1398
Db	1335	QTIAGKLYFASD-GKQVKSQFVT-YNGKVHYHADSGELQVNRFEADKGNWYLYDSNG	1392
QY	1399	NLVKGFQEIQDKWTHFBEVYTGQAKGAALVNGQOLYFDVDSGTVQKGDFTWD	1450
Db	1393	EALTQSQR-----NDQRFVYFQ-REGQVKGQDVAID	1422

```

; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

```

RESULT 3
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper

Db 924 KVMADVPDMYAFPEKEVWVTRVVDKYGTPVAGSQIKNTLYVVDKSSGKQQAQYGA 983
QY 990 FLDELKAKYPMFTNMISTGRPIDPSTKIKOWEAKYFNGTNTVLGKAGYVLSDDATGY 1049
Db 984 FLEELQAKYPELFAKQISTGVPMDBSVKIKQWAKYFNGTNTVLGKAGYVLSDDATGY 1043
QY 1050 FTVNENG--FLPASFTGONAKTGYDGTGMAYYSTSGNKAVNSFIYEGGHYFFDKD 1107
Db 1044 FNISDNKEINFLPKTLL--NQDSQVGSYDQKGYVYSTSGYQAKNTFISEGDKYFFDN 1102
QY 1108 GHMVTGSYKAEENDYFELPNGIQMRDAIYQDAQNSYYVGRGTILYKGDNVPFVDPNN 1167
Db 1103 GWTGCA--QINGVNYIFLSNGLQLDALKNEDGYIYNGDGRYE--NGYQFM---- 1156
QY 1168 ANKTVRYPDANNVMAIGYRNMYGQTYYPDENGFOAKGOLLTDDKG--THYFDEEDNGAMAK 1226
Db 1157 --SGVWRHEN--NGEMSVGLTVIDQVQYFDENGYQAKGFVTTADGKIRYFDKQSGNMYR 1213
QY 1227 NKFEV--NVGDDWYMDGNGNAVKQYPPVNNQILYFNPETGVQVKGQFITDAQGRTSYYDAN 1285
Db 1214 NRPFIENEKGWLYLGBDGAATGSGQTINGQHLIFR--ANGVQVKGFEVTDHGRISYYDGN 1272
QY 1286 SGALKSSGFFTPNGSDWYVAENGYYVKGFKQVAENQDQWYFDTTGGKQAKGAAGVDRD 1345
Db 1273 SGOQINRP-----VRNAQGWYFED--NNGYAVTGARTINGQL 1309
QY 1346 LYFNPDSGVQVKGDFATDSGNTSFYHGDNKGVKVGFFTTGNNAWYADNNGNLVKGFQ 1405
Db 1310 LYFRA--NGVQVKGFEVTDYGRISYYDGNSGQIRNRFRVNAQGWYFEDNNGYAV-- 1364
QY 1406 ELDGKWHYFDEVTGQAKGAALVNGQOLYFDVDSGLQVKGDFVTDGQNTSYDYNVSGDK 1465
Db 1365 -----TGARTINGQHLIFRA--NGVQVKGFEVTDHGRISYYDGNSGDQ 1406
QY 1466 KVNQFFTTGDNWYADGQNLAKGRKSTDNDLYPDATGQVKGQOLYSID--GRNYIFD 1524
Db 1407 INRFRVNAQGWYFEDNNGYAVTGARTINGQHLIF--RANGVQVKGFEVTDYGRISYYD 1465
QY 1525 SSGSNMKA 1532
Db 1466 ANSGERV 1473

RESULT 6

US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans

US-09-740-274-2

Query Match 41.7%; Score 3434.5; DB 2; Length 1475;
Best Local Similarity 46.6%; Pred. No. 9.1e-200;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;
QY 1 MEKLLHYKLHKYKGGHWVIAVAS--IGLYSLVCAGTVAEDKAVNDTTAAQTAVGVDTGQDQ 59
Db 1 MDKRYRYKLRYKRGKGVTVSVASAVMTLTLSGGLVKAD---SNESKSO----- 46
QY 60 ATTNDANTNTTDDTADQSAN-----TNODQAGSQSNNOQAKQDQTA----- 103
Db 47 -ISNDSNTSVV---TANESNVITEATSKQEAASSQTNHTVTTSSTSVVNPKEVSNP 102
QY 104 -TDRNQANSQTDNNQAATDQATS PATDGTFSVORRDAANVATAADQEGQTPAPEQEKAAL 162
Db 103 YTVGETASNGEKLQNTT-----TVDKTS---EAAANNISKQTTEADTDVIDDSSNAANL 153
QY 163 ----SLDNVVKLIDGKYVYVQADGSYKKNFAITVNGQMLYFSDTGCALSSTSYFSQGT 217
Db 154 QILEKLPNVKEIDGKYVYVYDNGKVRTNFTLTADGKILHFD--ETGAYTDTSDITVNOIV 212
QY 218 TNLVDDFSSHNAKAYDSTAKSFELVNGYLITANSWYRPAGILRNGQTWAEASNENDLRPLMS 277
Db 213 TTRSRLYKYNQVYDRSAQSFEHVDHYLTAEWYRPKYLKDGKTWTQSTQTEKDFPLMT 272
QY 278 WMPDKDTQVAYVYNNKYLANSNETEVTNETSQVDLNKEAQSIOTKIEQKITSNDSTQWL 337
Db 273 WMPDQETQRYVYNNQAOLGINKT--YDDTSNQLNLNIAAATQAKIEAKITTLKNTDML 331
QY 338 TAMEAFVAAQPKWNNSTEN--FNKGDHLOGGALLYTN--SDLTTPWANSDYELLNRTPOQD 395
Db 332 QTISAFVKTQSAWNSDSEKPF--DHLQNGAVLYDNEGKLTTPYANSVRYLNRTPNTQGT 389
QY 396 TK--KYFTEGGGGYEFLLSDNDVNSNPVQAEQLNQLHYLANWGDI VNGDXDANPDGVR 453
Db 390 KDPRTYADNTIGGYEFLLANDVNSNPVQAEQLNQLHYLANWGDI VNGDXDANPDGVR 449
QY 454 VDAVDNVADLLQVYSNYFKNYKVTYDSEANALAHISILEAWSLNDNQNETNGTALSI 513
Db 450 VDAVDNVADLLQIAGDYLKAAGYKHKNDKAAANDHLSILEAWSNDNTPYLHDDGDNNMN 509
QY 514 DNSSRLTSLAVLTQPGQRIDLSNLSSESVNKERANDTAYGDTIPTYSFVRAHDSVQTV 573
Db 510 DNKRLSLFLSLAKPLNQRSGMPLITNSL--VNR2DDNABTAAPVSPYSIRAHDSVQDL 568
QY 574 IAKIVKEIDTNSDGYTFTLDQDAFKIYNEDMAKVNKYTHYTHYNI PAAYALLLSNMSV 633
Db 569 IADIIKAEINPNVGYSTFMEIKKAFELYNKDLATEKKYTHYNTALSYALLTNKSSV 628
QY 634 PRVYDGLYTDQVMAKSPYDIAIATMLOGRIAYVGGQSEEVHKNVNGNQILSSVRY 693
Db 629 PRVYDGMFTDDQVMAKHTINYEAIETLLKARIKYVSGQAMRNQOV--GNSIITSVRY 687
QY 694 GQDMSADDTQGTDLRSRTSGLVTLVSNQPNLDL--GGDSLTVNMGRAHQAQVRPLILGK 752
Db 688 KGKALKATDT--GDRTRTSGVAVIEGNNPSSLKASDRVVVNVNGAAHQRQVRPLLLTDD 746
QY 753 DGVSQYLKQSD--TNIVKYVTDANGNLTFTADDIKGYSTVDMSGVLA VVPVGAK--DGQDV 809
Db 747 NGIKAYHSDQEAAGLVRYTNDREGELIFTAADIKGANPOVSGYLVGVVPGAALIKWFAL 806
QY 810 RVAADTNQADGKSLKTSAAALDSQVIYEGFSNPQFANNADYTNKXIAENADFFKKLGI 869
Db 807 RLARPHQQMA---SVHQNAALDSRVMEGFSNFQAFATKKEEYTNVVI AKNVDFKPAEWGV 863
QY 870 TSFEMAPQVYSATDGSFLDSIIONGYAFSDRVYDLMSKNKYGSKDDLANKALHANGI 929
Db 864 TDFEMAPQVYSSTDGSFLDSVIONGYAFSDRVYDLMSKNKYGSKDDLANKALHANGI 923
QY 930 QAIADWVPDQIYQLEPGEVWVTA RYNSYGNPTFDAYINNALYATNTPKSSGSDYQAOYGA 989
Db 924 KVMADVPDMYAFPEKEVWVTRVVDKYGTPVAGSQIKNTLYVVDKSSGKQQAQYGA 983

990 FLELAKAKYPMFTVNMISTGKPIDBSTKIKOWEAKYFNGTNVLGKAGYVLSDDATGKY 1049
1049 FLELAKAKYPMFTVNMISTGKPIDBSTKIKOWEAKYFNGTNVLGKAGYVLSDDATGKY 1049
984 FLEELQAKYPELFAKQISTGVPMPSVKIKOWSAKYFNGTNILGRGAGYVLDKQATNTY 1043
1050 FTVNENG--FLPASTGDNQAKTYGTYDGTGMAYYSTSGNKAIVNGFIYEGHYFFDKD 1107
1044 FNISDNKEINFLPKTL--NODSQVGSYDGKGYVYYSYGQAKNTFFISEGDKWYFFDN 1102
1108 GHMVTGSYKAEDGNDYFFLNGIQMRDAIYQDAQNSYVYGRGTYLYKGDNNYFFVDPNN 1167
1103 GYMTGA--QINGVNYFFLSNGLQRLDALIKNEDGTIAYYGNDRYE--NGYQFN---- 1156
1168 AKNTVRYFDANNVMAIGYRNNMYGTYFFDENGFOAKQLLTDKKG--THYFDEDNGAMAK 1226
1157 --SGVWRHFN--NGEMSGLTVIDGQVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYR 1213
1227 NKFPV--NVGDDYMDGNGNAVKQYFVNNQILYFNPETGVQVKGQFIDTAQGTSYDAN 1285
1214 NRFIENEBGKWLGLGEGDGAATGSGQTINGOHLIFR--ANGVQVKGFEVTDHGRISYDGN 1272
1286 SGALKSSGFFTPNGSDWYAEYVYKGFQKVAENODQWYFDDTTGKQAKGAAYDGRD 1345
1273 SGDOIRNRF-----VNAQGFYFD--NNGYAVTGARTINGQ 1309
1346 LYFNPDSGVQVKGDFATDESNTSYFHDGNDGKVGGFPTTGNNAWYADNNGNLVKGFP 1405
1310 LYFRA--NGVQVKGFEVTDYGRISYVYDNGSGQIRNRFVNAQGFYFDNNGYAV---- 1364
1406 EIDGKWHYHDEVTGQAKGAALVNGQQLYFDVDSGLQVKGDFVTDGQNTSYVDVNSGDK 1465
1365 -----TGARTINGOHLIFRA--NGVQVKGFEVTDHGRISYVYDNGSGDQ 1406
1466 KVGPFPTTGDNAWYADGQNLAKGRKSIDNODLYFDPATGKQVKGQLYSID--GRNYFD 1524
1407 INRFRVNAQGFVFDNNGYAVTGARTINGOHLIF--RANGVQVKGFEVTDYGRISYVD 1465
1525 SGSGNMAK 1532
1466 ANSGERVR 1473

RESULT 7

US-08-793-824-2
Sequence 2, Application US/08793824
Patent No. 5981838
GENERAL INFORMATION:
APPLICANT: Simpson, Christine Lynn
APPLICANT: Giffard, Philip Morrison
APPLICANT: Jacques, Nicholas Anthony
TITLE OF INVENTION: Genetic Manipulation of Plants to
TITLE OF INVENTION: Increase Stored Carbohydrates
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: Griffith Hack & Co
STREET: Level 8, 168 Walker Street
CITY: No. 5981838th Sydney
STATE: New South Wales
COUNTRY: Australia
ZIP: 2060
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,824
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM7643
FILING DATE: 24-AUG-1994
TELECOMMUNICATION INFORMATION:

TELEPHONE: 61 2 9957 5944
TELEFAX: 61 2 957 6288
TELEX: 26547
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1577 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Streptococcus salivarius
US-08-793-824-2

Query Match 41.6%; Score 3429.5; DB 1; Length 1577;
Best Local Similarity 44.7%; Pred. No. 2e-199;
Matches 738; Conservative 250; Mismatches 471; Indels 193; Gaps 40;

QY 1 MEKKLHYKLHKVGHVWTIAVASIGLVSLVGAGTVAEDKVAANDTTA----- 47
Db 1 MENKVRFKLHKVKNWVTIGVTTLSMVALAG--GSLLAQGVKVEADTSAPNGDGLQLSLED 59
QY 48 -----QATGVDTGQDQATNDANTNTTDTADQASANTNQDQSDGNNODQAKQDT 101
Db 60 GTASLVTITTTTEQASQAQASAVATASVSHETSFOATSQVSEATAQATSPVASQEV 119
QY 102 ANTDNRQADNSOTD-----NNOATDQATSPATDGTSVQRD-----AANVAT-A 144
Db 120 AVSSQTSQSGQETQTEQVSGQSTQVAGQTSQASTPSVTBOARPRVLTNAAPATRA 179
QY 145 AD-----OBGO- 150
Db 180 ADSTIRINANRNTNITITAGTTPNVTITGPNTPKPNVTVTSPNCTRPNVTITVQNPQ 239
QY 151 -----TAPSEOEKSA-----ALSLD-----NVKLIDGKYVYQVADGSGYKKNFAITVNGQM 195
Db 240 NKPVPQSFQSPQNPQNPQNPQNPQNPQNPQNPQNPQNPQNPQNPQNPQNPQNPQNPQ 298
QY 196 LYFSDSDTGAL--SSTSTYSFSGTTLNVDVDFSSHNAKAYDSTAKSPELVNGYLTANSWYRP 253
Db 299 YYFD--ETGAVDQSKPLRADAI PNNSI--YAVYQAYDTSSKSPFHLNFLTADSWYRP 355
QY 254 AGILRNGQTWEASNENDLRPVLMSWMPDKDTQVAVYVNMKNKYLSEANETVETNETSOVDLN 313
Db 356 KQLADGKNWTASTKDYRPLMTWPKVQVQVYVNLNYSQGFNGKNTTDTDMWS-YDLA 414
QY 314 KEAQSIQTKIEOKITSDNSTOWLRTAMEAFVAAQPKWNMSTE--NFKVG--DHLOGGALLYT 371
Db 415 AAAETVQRGIBERIGREGNTTWLRQLMSDFIKTQPGWNSSESDNLLVGKDLHQLGGALTFL 474
QY 372 NSDLTPWANSQYRLNRTPTQODGTYKYTEGEGGYEFLLSNDVDNSNPVQAEOLNQL 431
Db 475 NNSATSHANSDFRLMNRPTTGTGRKYHIDRSNGGYELLANDIDNSNPAVQAEOLNWL 534
QY 432 HYLNNWGDIVMGDKDANFDGVVDAVDNVDNADLLOVSNYFKDNYKVTSEANALAHISI 491
Db 535 HYIMNIGSILGNDPSANFDGVRIDAVNDVADLLOIADSYFKEKYRVADNEANALAHISI 594
QY 492 LEAMSLNDQXNEDTNGTALSIDNSSRLTSVLAVLTKPQQRIDLNLISVSYNKERANDT 551
Db 595 LEAMSYNDHQNKDTKGAQLSIDNPLRETLTTLFLRKSNGYRGLSERVITNSLN-NRSSSQ 653
QY 552 AYGDITPTYSFVRADHSEVOTVIAKIVKEKIDTNSDGYTFTLDQLKDAFKIYNEDMAKN 611
Db 654 KHTPRDANYIFVRAHSEVQAVLANIIKQINPKPTDGTFTTMDLQKFAEYINADIAKAD 713
QY 612 KTYTHYNI PAAYALLSNKSVPRVYVYGDLYTDDQYMAKGSPPYDAATATMLQGRYAVYS 671
Db 714 KKYTOYNI PAAYATMLTNKDSITRYYVYGDLYTDDQYMAKGSPPYDAATATMLQGRYAVYS 773
QY 672 GQSEEVHKVNGNNQILSSVRYGQDLMGADDTQGTDLSTGLVTLVSNDPNLDLG-GDS 730
Db 774 GQDMKVTKLG--YEIMSSVRYKGAEEANQL-GTAEITRNGQMLVLTANRPMKLGANDR 831

```
QY 731 LTVNMGRAHANOAYRPLILGTXDGVQSYLKSDS--TNIVKYTDANGNLFTTADDIKGYST 788
Db 832 LVVNGAAHKNQAYRPLLSKSTGLATYLDKSDVPAGLVRYTDNQNLTFTTADDIAGHST 891
QY 789 VDMGVLAVVVPVGAQGDVVAADTQKADKSLKTSAAALDSQVIVYEGFNFQDFANN 848
Db 892 VEVSGYLAVVVPVGAQGDVVAADTQKADKSLKTSAAALDSQVIVYEGFNFQDFVKT 950
QY 849 DADYTNKXIAENADFFKLGITSFEMAPQVVSATDGSFLDSIIQNGYAFSDRYDLAMSKN 908
Db 951 PSQYTNRVIAQNAKLFKEWGITSEFEAPQVVSQDGTFLDSIIENGAFEDRYDIAMSKN 1010
QY 909 NKYSKDDLALNALKALHANGIOAIAADVPDQIYQLPGEVVTAKTNSYGNTPPAYINN 968
Db 1011 NKYSGLKOLMDALRALHAEIGISAIADVPDQIYNLPGEVVTASRTNSYGTGPRNAEYIN 1070
QY 969 ALYATNTKSSGSDYOAGVGAFLDELKAKYPMFTVMNISTGKPDPSKIKQWEAKYFN 1028
Db 1071 SLUYAAKTRTFGNDFQKYGGAFLDELKAKYPAIFERVQISNGRKLTTNEKITQWSAKYFN 1130
QY 1029 GTNVLGKAGYVLSDDATGKYFTVNENGDPLPASFTGQNAKTGFYDGTGMAYYSTGN 1088
Db 1131 GSNIGTGARYVLDNATNQYFSAKAGQTFLEPKMT--EITGSGFRFVGVDDVQVLSIGY 1188
QY 1089 KAVNSFIYEGGH--YYFDKGDHMTGYSKAEKNDYIFLPLNGIQMRDAIYQDAQNSYY 1147
Db 1189 LAKNTFIOVGANQWYFYDKNGNMVTGE--QVIDGKYFFLDNGLQLRHLVROGSDGHVYY 1247
QY 1148 GRTGI--LYKGDWNPVDPNNANKTVFVFDANNVMAIGYRNMVGYTYFDS--NGFQAKG 1205
Db 1248 DPKGVQAFNG--FYDFAGP--ROVRYFDNGQMYRGLHDMYGTTFYFDEKTLQAKD 1301
QY 1206 QLLT--DDKGTHTYFEDENGAAMKKNFVNVGDD--WYMDGNGNAVKGQVYVNNQILYFNP 1262
Db 1302 KFIRPADGRTRYFIPDTGNLAVNRPAQNPENKAWYLDNSGYAVTGLQTINGKQYFDE 1361
QY 1263 TGQVQKQFIPDAQRTSYIDANSALKSSGFFTPNGSDWYAEANGYVYKGFQVAENQD 1322
Db 1362 -GRQVKGHEVT--INNQRVFLDGSGEIAPSRFVTEN-----N 1396
QY 1323 QWYVFDQTTGQAKGAQVDRDLVFNPDGSGVQVKGDPATDESNGTSFYHGDNGDKVVG 1382
Db 1397 KWIYVD--GNGKLVGAQVINGNHYYFNNDYS--QVKGAWA-----NGRYDGDGSGQAVSQ 1449
QY 1383 FTTTGNNAWYADNNGNLKGFQEIWKYHFDVGTGQQAAGALVNGQQLYFDVDSGIG 1442
Db 1450 FIQIAANQWYLNQDGHKVTGLQINNKVYFGE-----SNGAQ 1487
QY 1443 VKGDFVTDGQNTSYIDVNSGDKKNGFFTTGDNWYADGQGNLAKGRKSIDNQDLYPD 1502
Db 1488 VRGKLLT--VQGGKCYFDAHTGEGVNNRFYEAARGCWYFNSAGQAVTGQOVINGKQLYPD 1546
QY 1503 PATGQVKGQVLSIDGRNYPYDPSGSGNMAKNR 1534
Db 1547 -GSGRQVKGRYVYVGGKRLFCDAKTGELQR 1577
```

RESULT 8

```
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
```

```

; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
; US-09-210-361-4

Query Match      38.4%; Score 3161.5; DB 2; Length 1375;
Best Local Similarity 48.5%; Pred. No. 3.1e-183;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;

QY 1 MEKKHLKHLKHKHWHVIAVAS--IGLVSLGA--GTVSADKVDNTTAQATVGVDTG 56
Db 1 MEKKVRFKLRKVKRWVTSIASAVVTLTSLGSLVKADSTDRQQAQVTEQASL----- 55
QY 57 QDQATTNDANTNT--TDDTDAQSANTNODQAG--SDQSNNOQAKODTANT-----DRNQ 108
Db 56 ---VTTSEAAKETLTATDTSTATSATSOPTATVUNVSTTNOSTTNTTANTANFVKPTTT 112
QY 109 ADNSQTDN-----NQATDQATS-----PATDGTSVQRRDAANVATAADQEG---QTAP 153
Db 113 SEQAKTNSDKIITTSKAVNRLTATGKVPANNNTAHPKVTVDKIVPIKPKIGLKQPS 172
QY 154 SEQEKSAAL--SLDNVKLIDGKYVYQADGSKYKKNFAITVNGQMLYFSDTGALSTSTYS 212
Db 173 LSQDDIAALGNVKNIRKVGYYKEDGTLQKNYALNKGTFEFD--ETGALSNNLTFS 231
QY 213 FSQGTIN--LVYDDFSSHNKAYDSTAKSPELVNGYLTANSWYRPAGLRNGQWTEASND 270
Db 232 KKGNTTNDNTNSFAQYNQVSTVDANFEVHDYHILTAESWYRPAKILKDKGTWTQSTEKD 291
QY 271 LRPVLMSWMPDKDTQVAYVYNNKYLSANETEVNTNETSQVDLNKEAQSITQKIEQKITS 330
Db 292 FRPLMTWPDQETQRYVYNNYNAQLGHTQVNT--ATSPQLNLAAQTIOTKIEKITAE 350
QY 331 NSTQWLRAMEAFVAAQPKWNNSTEN--FNKGHDHLOGGALLYT--NSDLTPWANSDFLLNR 388
Db 351 KNTNWLRTISAFVKTQSAWNSDSEKPF--DHLQGALLYSNNSKLTSAQNSRYRLNR 408
QY 389 TPTQODGTG--KYPTGEGEGGYEFLLSNDVNSNPVQAEQLNLHLYLMNWGDIVMGDKD 446
Db 409 TPTNGTGGKDPRTADRTIGGYEFLFLANDVNSNPVQAEQLNLHLYLMNWGDIYANDPD 468
QY 447 ANFDGVRVDAVDNVDNADLLQVYNSYFNKDYKVTDSANALAHISILEAWSLNDNQYNEPT 506
Db 469 ANFDSIRVADVNDVDADLLQIAGDYKAAKGTHKNDKAAANDHLSILEAWSYNDTPYLHD 528
QY 507 NGTALSIDNSRLTSVLAVITKQPGQRIDLSNLSISVKNKERANDTAYGDTITPTYSFVRAH 566
Db 529 GDNMINMDNRLRLSLYSLAKPLNQRSGMNPITNSL--VNRITDDNAETAAPVSYSFIRAH 597
QY 567 DSEVQTVIAKIVKEKIDTNSDGYTFTDQLDKAFKIYNEDMAKVNKYTHYNYPAAYALL 626
Db 588 DSEVQDLIRNIIRTEINPNVVGYSFTTEIKAKAFIYNKDLATATEKYTHYNTALSYALL 647
QY 627 LSNMESVPRVYVYDLYDDGQYMAKSPYDAIATMLQRIAYVSGGQSEEVHKVNGNNO 686
Db 648 LTKSSVPRVYVYDMDFTDDGQYMAHKTINYEAETLLKARIKYVSGGQAMRNOQV--GNSE 706
QY 687 ILSSVRYGODLMSADDTQGTDLRSRTSLVTLVSNDPNLDL--CGDSLITVNGRAHQAQYR 745
Db 707 IITSVRYGKGALKATDT--GDRTRTSGVAVIEGNPNLSLKLKASDRVVNVNNGAAHKNQYR 765
QY 746 PLILGTGKGVSQSYLKSDS--TNIVKYTDANGNLFTTADDIKGYSTVDMSGYLAWVPVGA 804
```


Db 766 PLLETTDNGIKAVHSQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLVWVPVGA 825
Qy 805 DGQDVRVAADTNOKADCKSLKTSAAALDSQVIYEGFSNFQDPANNDADYTNKKIAENADFF 864
Db 826 ADQDVRVAASTAFSTDGKSHQNAALDSRVWVFSGFSNFQAFATKBEYTNVIAKNVDF 885
Qy 865 KKLGITSFEMAPQVYATDGSFLDSIIQNGYAFSDRYDLAMSNNKYGSKDDLALAKAL 924
Db 886 AEWGVTDFEMAPQVYATDGSFLDSIIQNGYAFSDRYDLAMSNNKYGSKDDLALAKAL 945
Qy 925 HANGIOALADWVPDQIYOLPGEVBTAKRTNSVGNPTFDAYINNALYATNTKSSGSDYQA 984
Db 946 HSKGIKVMADWVPDQIYOLPGEVBTAKRTNSVGNPTFDAYINNALYATNTKSSGSDYQA 1005
Qy 985 KYGGAFLDELKAKYPMFTNMISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDD 1044
Db 1006 KYGGAFLDELKAKYPMFTNMISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDD 1065
Qy 1045 ATQKTYTVMENGDFLPASFT----GQONAKTGPYDGTGMAYYSTSGNKAVNSFIYEGGH 1100
Db 1066 ATTYFSLVSDNTFLPKSLVNPNGTSSVTGLVFDGKGYVYSTSGNKAQNAFISLGN 1125
Qy 1101 YYYFDKGHMVTSYKAEQNDYVFLPNGIQMRDAIYQDAQNSYYVYGRGILYK----- 1155
Db 1126 WYFFDNGYVMTGA-QSINGANYFLSNGIQLNAYIDNGNKVLSYVYGNDRYENGYYL 1184
Qy 1156 -GDMWYFPVDPNANKTVFRYPDANKVMAIGYRNMYGQTYFDPENGFOAKGOLLTDKG- 1213
Db 1185 FGQW-----RYFO-NGIMAVGLTRVHGAQVYFDASFOAKGQFITTADGK 1229
Qy 1214 THYFEDNGAMAKNFV-NVGDWYVMDGNGNAVKGOYPVNNQILYFNPETGVQVKGQFT 1272
Db 1230 LRYFDRDSGNQISNRVFRNSKGEWFLFDHNGVAVTGTVTFGQRLYFKP-NGVQAKGEFI 1288
Qy 1273 TDAQRTSYDANSALKSSGFTFPGSDWY-YAENGYVYKGFQV 1317
Db 1289 RDANGYLRYDPSNGNEVRNFRVNSKGEWFLFDHNGIAVTGARVV 1334

RESULT 9
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match { 38.4%; Score 3161.5; DB 2; Length 1375;

Best Local Similarity 48.5%; Pred. No. 3.1e-183;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;
Qy 1 MEKKLHVKKLVKHWVTIAVAS--IGLVSLVGA--GTVAEDKAVANDTTTAQATVGVDTG 56
Db 1 MEKKVPFKLVKVKRWVTVSIASAVTLTSLSSGLVKADSTDDRQQAQVTSQASL----- 55
Qy 57 QDQATTNDANTNT-TDTDADQSANNTNQDQAG-SDQSNNOQDQAKQDPTANT-----DRNQ 108
Db 56 ---VTTSEAAKETLTATDTSTATSATSQPTATVTDNVSTTNTTNTTANTANFVVKPTTT 112
Qy 109 ANDSQTDN-----NQATDQATS-----PATDGTSVQRDAANVATAADQEG---QTAP 153
Db 113 SEQAKTNDSDIKITTSKAVNRLTATGKFPANNNTAHPKTVTDKIPIKIGLQKQPS 172
Qy 154 SEQEKAAL-SLDNVKLIIDGKYVVOADGSKKNFALITVNGOMLYPDSDTGALSSSTSTYS 212
Db 173 LSQDDITAAAGNVKIRKNGKYYKEDGTQKNYALNNGKTFPFD-ETGALSNNLTLP 231
Qy 213 FSQGTN--LVDDFFSHNKAYDSTAKSFELVNGYLTANSWYRPAGLRNGQTWEASNEND 270
Db 232 KKGNIITNDNTNSFAQYNQVYTDVANFEHVDHYLTAESWYRPKYILKDGKTTWTQSTEKD 291
Qy 271 LRPVLSWVPDKDTQVAYVYNNKYLISANETEVNETSQVDLKNKAQSIOTKIEQKITS 330
Db 292 FRPLMTWVPDQETQRYVYNNQAQIGIHQTYNT-ATSPQLNLAAQTIQTKIEEKITAE 350
Qy 331 NSTOMLRTAMEAFVAAQPKNNKSTEN-FNKGDHLQGGALLYT--NSDLTPWANSDYRLNLR 388
Db 351 KNTNWLRLQITSAFVKTSQSAWNSDSEKFPD--DHLQGLLYSNNKSLTQANSNRYILNR 408
Qy 389 TPTQDQGT--KYFTGEGGEGYEFLLSNDVDSNPVVAEQNLNQLHLMWGDIVMGDKD 446
Db 409 TPTNQTKGKPRYADRTTIGYEFLLANDVDSNPVVAEQNLNQLHLMWGFNIYANDPD 468
Qy 447 ANFDGVRVDAVNADLLQVTSNTPKQNVKVTDSANALAHISILEASLNDNQVNETD 506
Db 469 ANFDSIRVDAVNDADLLQIAGDYLKAAKGHIKNDKAANDHLSILEASYNDDTPLYHDD 528
Qy 507 NGTALSIDSRLTSLAVLTQPGQRI DLSNLISSEVKNKERANDTAYGTIPTYSFVRAH 566
Db 529 GDNMNMNRLSLLSLYSLAKPLNQSGNPLITNSL-VNRTDDNAETAAVPSYFIRAH 587
Qy 567 DSEVQTVIAKIVKEIDTNSDGYTFLDQKDAFKIYNEDMAKVNKTYTHYNIPAAAYALL 626
Db 588 DSEVQDLIRNIIRTEINPNVGVYFTTEIKKAFIYNKDLATEKKYTHYNIALSVALL 647
Qy 627 LSNMESVPRVYGDLYTDDGQYMAKSPYDAIATMLQGRIAYVSGGQSEEVHKNVGNQ 686
Db 648 LTNKSSVPRVYGDLYTDDGQYMAKHTINYEAIETLLKARIKYVSGGQAMRNQV-GNSE 706
Qy 687 ILSSVRYGQDLSADDTCGTDL-SRTSGLVTLVSNQDNLDL-GGDSLTVMNGRAHQAAYR 745
Db 707 IITSVRYGKALKATDT-GDRTTSTGVAIVIEGNPNSRLKASDRVVNNGAAHQAAYR 765
Qy 746 PLILGTGQVQSYLKDSQ--TNIYKTYDANGNLTFTADDIKGYSTVDMSGYLAWVVPVGA 804
Db 766 PLLLTTDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLVWVPVGA 825
Qy 805 DGQDVRVAADTNOKADCKSLKTSAAALDSQVIYEGFSNFQDPANNDADYTNKKIAENADFF 864
Db 826 ADQDVRVAASTAFSTDGKSHQNAALDSRVWVFSGFSNFQAFATKBEYTNVIAKNVDF 885
Qy 865 KKLGITSFEMAPQVYATDGSFLDSIIQNGYAFSDRYDLAMSNNKYGSKDDLALAKAL 924
Db 886 AEWGVTDFEMAPQVYATDGSFLDSIIQNGYAFSDRYDLAMSNNKYGSKDDLALAKAL 945
Qy 925 HANGIOALADWVPDQIYOLPGEVBTAKRTNSVGNPTFDAYINNALYATNTKSSGSDYQA 984
Db 946 HSKGIKVMADWVPDQIYOLPGEVBTAKRTNSVGNPTFDAYINNALYATNTKSSGSDYQA 1005
Qy 985 QYGGAFDLDELKAKYPMFTNMISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDD 1044

Db 1006 KYGAFLLELOAKYPELPARKQISTGVPMDPSVKIKQWAKYENGNTNILGRGAGYVLKQD 1065
QY 1045 ATGKYFTVNEGDFLPASFT---GDONAKTGYDGTGMAYYSTSGNKAIVSYFIEGGH 1100
Db 1066 ATWYPSLVSDNTFLPKSLVNPNGHTSSVTGLVDFGKGVYVYSTSGNQAQNAFISLGN 1125
QY 1101 YYYFDKDHMTGSKYAEADGNDYFLPNGIQMRDAIYQDAQNSYYVYGTGLYK----- 1155
Db 1126 WYFDNNGYMTGA-QSINGANYPLSNGIQLRNLAYDNGNKVLSYVYDNGRRYNGYYL 1184
QY 1156 -GDMWYFPFDPNNANKTVFRYFDANNVMAIGYRNMVYGTYYFDENGFOAKGOLLTDDKG- 1213
Db 1185 FGQW-----RYFO-NGIMAVGLTRVHAGVQYFDFASGFOAKGQFIITADCK 1229
QY 1214 THYFDEEDNGAMAKNFV-NVGDDWYMDGNGNAVGQYPVNNQILYFPNPETGVQVKGQFI 1272
Db 1230 LRYFDRDNGNQISNRVPRNSKGWFLFDHNGVAVTGTVTFNGQRLYFKE-NGVQAKGEFI 1288
QY 1273 TDAQGRTSYDANSALKSSGFPFNGSDWY-YAENGYYVYKGFQV 1317
Db 1289 RDANGYLRYYDPNSGNEVRNFRVNSKGWFLFDHNGIATVGARV 1334

RESULT 10

US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match 29.4%; Score 2422; DB 2; Length 2057;
Best Local Similarity 37.3%; Pred. No. 4.1e-138;
Matches 649; Conservative 220; Mismatches 577; Indels 294; Gaps 64;

QY 43 NDTTAQATVGVDTGQDQATNDANTTTDT-DADQSANNTN-----QDQAGSD 89
Db 210 DSVTGKASNDVIVNGKAQGYDAQGNLKKSVYVADSSGQTYYPDNGQPLIGLQTTIDGNL 269
QY 90 QSNQD--QAK--QDANTDRNOADNSOTDNNQATDQATSPATDGTQSVQRDAANVATA 144
Db 270 QYFNQGGVQVKGQFQVNNKRIYFAPN--TGNNAVANTEIINGKLQORDANGQVKN-APS 326
QY 145 ADQEGTAPSEQFSKSAALSDNVKLDGKYVYVQADSGYKQNFPAITVNCQMLYFDSDTGA 204
Db 327 KDVAAGTFFD---ANGVMTGLQITISGKTYLDEQHLRKNVAGTFNNQFMYFVADTGA 383
QY 205 LSTSTSYFSQGTNLVDFSSH--KAYDSTAKSPFELVNGYLTANSWYRPAGILNGOT 262
Db 384 GTAIBEQDQGLVQSNTPHNAKSYDKS--SPENVGGLTADTWYRPTDILKNGDT 441
QY 263 WEASNENLRLPVLMSWPKDQTVAYVYVNMKYL SANETVNETSQVDLNKEAQSIQTK 322
Db 442 WTASTEDRPLMTWPKQTOANYLNFMSSKGLGITTYTAQTSQKTLNDAAPIQTA 501
QY 323 IEQKITSNDSOWLRTAMEAFVAAQPKWNSTEN--FNKGDHLQGGALLY-TNSDLTPWA 379
Db 502 IEQKISLKSTLEWLRDAIDSFVKTOANWKNQTEDEAFDGLQWLQGGFLAYQDDSHRTPNT 561
QY 380 NS-DYELLNRTFPQDGTKKYFTTEGEGGYEFLLSNDVNSPNVQAEQLNLHLMWNG 438

Db 562 DSGNRKLRQRPINIDGSKD--TTDGKS-EFLANDIDNSNPIVOAEQLNLHLMWNG 618
QY 439 DIVMGDKDANFDGVRVDAVDVNNADLLQVSYNFKDYKVTDSSEANALAHISILEAWSLN 498
Db 619 SITGNDNANFDGIRVDAVDVNDALLKIAGDYFKALYGTDKSDANANKHLSILEDWNGK 678
QY 499 DNQYNEDTNGTALSIDNSRLTSLAVLTKQPCQRIIDL-----SNLISESVNKR 547
Db 679 DPQYVNOQNAQLTWDYTVTSQFNSLTHGANNRSMNYFLDTGYLNGDLNKKIVDKNR 738
QY 548 ANDTAY-----GDT--IPTYSFVRAHDSVQTVIAK-----IVKEKIDTNSDGYT 592
Db 739 PMSGTLVNRANSQDKVIPNYSFVRAHDYDQADPIRKAMIDHGIKMQD-----TFT 792
QY 593 LDQLKDAFKIYNEEDMAKVN--KTYTHYINIPAAAYALLSNMESVPRVYGDLYTDDQYMA 650
Db 793 FDLAQGMFEFYKQDENPSPGFKKYNDYNLPSAYAMLLTNKDTVPRVYGDMLYEGQYME 852
QY 651 KSPYPIDALATMLQRIAYVSGQ-----SEVHKVNGNQLLSVRYQODLMSADDTG 705
Db 853 KGTIYNPVISALLKARIKYVSGQTMATDSSGKDLKOGETDLTTSVRFKGIMTSDQTTT 912
QY 706 TDLS---RTSGLATLVSDNPNLDLGGD-SLTVNMGRAHANAQYRPLILGTGKGVQSYLKD 761
Db 913 QNSQDYKNQGGIGVIVGNPNDLKANDKITILHMGAKHNQYRALVLSNDSGIDVYDSD 972
QY 762 SDTNIYKVTANGNL-----TFTADD-----IKYSTVDMSGYLAVVVPGAKGDQD 808
Db 973 DKAPTLL-TNDNGDLIPHKTNTFVKDGTIINVEKSGSLNALISGLVGVVPGVSDSD 1031
QY 809 VR-VAADTNQKADGKSLKTSALDSQVIYEGSNFQDPFANNADYTNKKIAENADPFKL 867
Db 1032 ARTVATESSSNDGVSFHSNAALDSNVIYEGSNFQAMPSTSEQSTNVVVIATKANLFEL 1091
QY 868 GITSFEMAPOYVSATDG-----SFLDSIIQCYAFSDRYDLAMSKNN-----KYGSKDD 916
Db 1092 GITSFELAPQYRSGSDTNYGMSFLDSFLNNGYAFTRDYDLGFNKGADGNPNTKYGTDD 1151
QY 917 LANALKALHANGIAIADWVPDQIYQLPGEVYVTAKTNTSYGNPTFPDAYINNALYATNK 976
Db 1152 LRNAIEALHKGQAIADWVPDQIYALPGKEVVTATRVDERGNQLKDTDFVNLVYVANTK 1211
QY 977 SSGSDYQAYGGAFLDELKAKYDPDMFTVNMISTGKPIDBTKIKOWEAKYFNGTVNLGK 1036
Db 1212 SSGVDYQAYGGEFLDKLREYPSLFKQNVSTGQPIDASTKIKOWSAKYMNGTNLHRG 1271
QY 1037 AGYVLSDDATGKYFTVNEGD--FLPASFTGDQNAKTGFYDGTGMAYYSTSGNKAIVNSFI 1095
Db 1272 AYYVLKDWATNOYFNAIKTNEVFLPLQLQ-NKDAQGTGIFSDASGVKYYSISGVOAKDTFI 1330
QY 1096 YEG-GHYTYFDKDHMTG-----SYKAEADGNDYFLPNGIQMRDAIYQDAQ 1142
Db 1331 EDGNGWYTFDKGYMVRSSQGENPIRTVETSVNTRNGN-YVFMNPGVELRKGFGTDNSG 1389
QY 1143 NSYTYGRTGILYKGNWTFVDPNNANKTVFRYPANNVMAIGYRNMVYGT--YFPENGEF 1201
Db 1390 NYVYFDQCKMVR--DKYINDANN-----FYHLNVDTGMSRGLFKFDSDTLQYFASNGV 1442
QY 1202 QAKGQLLTDKGT-HYFEDDNGAMAKNFVNVGDDWYV-----1239
Db 1443 QIKDSYAKDSKGNKYFDSATGNDTQKATQWDGNGYIITDSDANNITGVNFDYTAIYT 1502
QY 1240 -----DGNNGNAVKGQYVNNQILYFPNPETGVQVKGQF-----1271
Db 1503 SSLREDGLPANAPYGVVTKDQNGNDLKWQY-INHTKQYEGQ--VQVTRQYTDKSGVSN 1559
QY 1272 -IT-----DAQRTSYDANSALKSSGFP-----PNSGDWYIAENGIVYKGFQV 1317
Db 1560 LITFAGGDLQGLRQWLD--SRALTWTPTFKTMNQISFISYANRNDGLFLNAPYQVKGY-QL 1616
QY 1318 AENQDQWYFDQTTGKQAKGAAGVGRDLVFNPDGSGVQVKGDFATDESNTSFYHGDNG 1377

Db 1617 AGHSNQ-YKCOQVT---IAGVANVSGKDWLSISFNGTQY---WIDSOALNTNTHDMNQK 1669
Qy 1378 KVVGGFFTTGN-----NAVY-----YADNNGNLVKGFBIDGKWHFDE---VTG 1419
Db 1670 VFN-----TTSNLDGLFLNAPYRQPGYKLAGLAKYNN-----QTVTVSQYFDDQGTWVS 1721
Qy 1420 QAKGAALVNGQQLYFDVDSGIQVKGDFVTDGQNTSYDYNVSGDKKVGNGFFTTGDNAWY 1479
Db 1722 Q-----VVLGGQTVVDNHALAQWQ---VSD---TDQQLVYVNSGRN-DGLFL---NAPY 1766
Qy 1480 YADG-----QGNLAKGRK-----SIDNQDLYFDPATGKQVKGOLVSI 1516
Db 1767 RGGSQSLIGHTADYNGHVQVTKQGDAYCAQWRLITLNNQQWVD---SRALSTTIMQA 1823
Qy 1517 DGRNYYPDS-----GSGNMAKNRFVRIGDQWI-YFGNDGGAATNL 1554
Db 1824 MNDNMVNSQRTDGLWLNAPYTMGAKWAGDTRSANGRYVHISKAYSNEVGNTYYLTNL 1883

RESULT 11
US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIORITY FILING DATE: 2000-06-28
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

Query Match 27.7%; Score 2285; DB 2; Length 1278;
Best Local Similarity 40.3%; Pred. No. 4.3e-130;
Matches 510; Conservative 189; Mismatches 409; Indels 156; Gaps 31;

Qy 52 GVDTCGQATTDANTNTTDDTADQSANTNQDQ-----AGSDQSNQD----- 95
Db 85 GDNVTDDQFSKNYA---TTGGNFDYKVGNGQVEFGSHATNQSDKDSQWIIIVLNGKE 141
Qy 96 -----QAKQDTANTDRNOA-----DNSQTDNNQATDQATSPAT---DGTSVQRDDA 138
Db 142 VKRQLVNDTKEGNAGFRNDVYKWPAINENSSMSGFGGI---ITLPVTYKQNVQLVHRFS 199
Qy 139 ANVATA-----ADQEGQTAP---SEQKSAALSIDNVLKIDGKYVYVQ-ADGSKYKKNFAIT 190
Db 200 NDVKTGEGNVDPWSELMPVKDSFQKNGPLKQFGLQTINGQQYIIDPTTGPQRKNFLQ 259
Qy 191 VNGQMLYFSDTCALSSSTYSFSQGTTLNLDVDFSHNKAYSTAKSFELVNGYLTANSW 250
Db 260 SGNNTIYFSDTGVGTGNLALQPAKGTVSSNEQYRNGNAAYSDDKSIENVNGYLTADTW 319
Qy 251 YRPAGILRNGOTWEASNDRLPVLMSWMPDKDTQVAYVYNNKY---LSANETEVNTE 307
Db 320 YRPQILKDGTTWDSKETDMDRILMWMPNTLUTQAYLYNNKQHGNLLPSALPFFNADA 379
Qy 308 SQVDLNKEAQSIQTKIEQKITSDNSTQWLRTAMEAFVAAQPKWNMSTENFN-KGDHLQGG 366
Db 380 DPAELNHYSEIQQNIIEKRISETGNTDMLRTLMHDFTVNNPMWKNDSNVNFSGIFQCG 439
Qy 367 ALLYTNSDLTWPANSYRLNLTPTQ-QDGTKYFTFEGGEGGYFELLSNDVNSNPVQA 425
Db 440 FLKYNSDLTPYANSYRLLRMPINIKQTYR-----GQEFLLANDIDNSNPVQA 491

Qy 426 EQLNQLHYLMNWGDIYVMGDKDANFDGVRVDAVDNVDNADLIQVYSNTFYKONYKYVTSDEANA 485
Db 492 EQLNWLYYLLNFCGTITANNDOQANFDSVRVADPNIDADLWNIADQYFNAAYGM-DSDAVS 550
Qy 486 LAHISILEAWSLNDNOYEDTNGTALTALSSRLTSLAVLTKQPGORIDLSNLISSEVVK 545
Db 551 NKHINILEDNWHADPEYFNKIGNPQLTMDDTIK-NSLNHGLSDATNRWGLDAIVHQSLS-A 608
Qy 546 ERANDTAYGDTPTYFSVRAHDSVOTVIKIDTNSDGYTFTLQDKDAFKLYNE 605
Db 609 DRENNSTENVIPDYSFVRAHDNNSOQIQNAIRDV--TGKDIHTFTFEDEQKIDAYIQ 666
Qy 606 DMAKNKTYTHYNIPAYALLSNMESVPRVYDGLYTDGQYMAKSPYDYDAIATMLQG 665
Db 667 DQNSTVKKNLYNIPASYAILLTNKDTIPRVYDGLYTDGQYMEHQTRYDYTLNLLKS 726
Qy 666 RTAYVSGGSESEVHKVNGKNNILSVRYQDLMSADDTQGTDLRSRTSLGLVTLVSDNPLD 725
Db 727 RVKYVAGGQSMQTMVSGGNNILTSVRYGKAMTATDT-GTDETRTQGGIGVVVSNTPNLK 785
Qy 726 LG-GDSLTVNMGRAHANOAYRPLILGTGQGVOSYKXSDTNI VKYTDANGNLFTADD-- 782
Db 786 LGVNDKVLHMGAAHKNQYRAAVLTITTDGVINTYSDQAP-VAMTDENGDLVLSHNLV 844
Qy 783 -----IKGYSTVDMSGYLAWVVPVGAQDQDVRVAADTNQKADGKSLKTSAAALDS 832
Db 845 VNGKEADTAVQYANPDVSGYLAWVVPVGAQDQDARTAPSTEKNSGNSAYETNAAFDS 904
Qy 833 QVIYEGFSNFQDPANNADYTNKKAENADPFKLGITSEFEMAPQVVSATDGSFLSIIQ 892
Db 905 NVIFEAFSNFVYTPTKESERANVRIAQNADFFASLGFTSFEMAPQVNSKDRFLDSTID 964
Qy 893 NGVAFSDRYDLAMSNNKYGSKDDLALNALKALHANGIOAIADWVPOIYOLPGEVVTAK 952
Db 965 NGYAFTRDYDLGMSPEPKYGTDEDLRNAIOALHKAQLQVMAWVPOIYNLPKKEVATVT 1024
Qy 953 RTNSYGNPTFDAYINNALLYATNTKSSGSDYQAQYGGAFDELKAKYPMFTVNMISTGKP 1012
Db 1025 RVDDRGNVWKDALINNNLYVNT-IGGGEYQKYGAGFLDKLQKLYPEITFKKQVSTGVA 1083
Qy 1013 IDPSTKIKOWEAKYFNGTNVKGAGYVLSDDATKYFTV-NENGDFLPASFTGD-ONAK 1070
Db 1084 IDPSQKITESAKYFNGTNILHRGSGYVLKADG-QOYVNLGTTTKQFLPQLTGEKKQGN 1142
Qy 1071 TGFYDGTGMAY-YSTSGNKAVNSFIYEG-GHYVYFDKDHVMTVGSYKADGNDYFPLN 1128
Db 1143 EGFVKGNDGNYFYDLAGNMVKNFTIEDSVGNWYFPDQDGKMWENKH----- 1189
Qy 1129 GIQMRDAIYQDAQGSYYGRTGILYKGDNWTVPFVDPNNANKTVFRYFDANNVMAIGYRN 1188
Db 1190 -----FVDVDS----- 1195
Qy 1189 MYGQ--TYVFDENGFOAKQLLTDKGTHTYFEDNGAMAKNFVNVDDMYMDGNNAV 1246
Db 1196 -YXEGTYFFLKNXGVSFRGLVQTDNGTYFD-NYGMVNRQNTINAGMIYTLDENGKLI 1253
Qy 1247 KQY 1250
Db 1254 KASY 1257

RESULT 12
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP

```

; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match      27.7%; Score 2285; DB 2; Length 1781;
Best Local Similarity 40.3%; Pred. No. 6.9e-130;
Matches 510; Conservative 189; Mismatches 409; Indels 156; Gaps 31;

QY 52 GVDTGQDQATTNDANTNTTDDTADQSANNDQ-----AGSDQSNQD----- 95
DB 588 GDNVTVDQFSKNYA---TTGNGFDYKVNQNGVFEFGWHATNQSDKDSQWIIVLVNGKE 644

QY 96 -----QAKQDQANTDRNQA-----DNSQTDNNQATDQATSPAT---DGTSVQRDA 138
DB 645 VKBQLVNDIKEGAAGFNNDVYKVNPAIENSSMSGFQGI--ITLPVTVKNENVQLVHRES 702

QY 139 ANVATA-----ADQEGQATP---SEQKSAALSIDNVKLDIGKYYVVO-ADGSYKKNFAIT 190
DB 703 NDVKTGEGNVDFPWSILMPVKDSFQKNGPLKQGLQTINGQOYYIDPTTGPQRKNFLQ 762

QY 191 VNGQMLYFSDTGALSTSYSPSQGTNLVDVDFSSHKNAYDSTAKSFELVNGVLTANSW 250
DB 763 SGNWLYFSDTGCVGTNALEQLQFAKGTSSNEQYRNGNAAYSDDKSIENVNGYLTADTW 822

QY 251 YRPAGILRNGQWEASNEENDLRVLMWMPDKDTQVAYVNMNKKY---LSANETEVTNET 307
DB 823 YRPQILKQGTWTDKSDTDMRILMVWHPNTLITQAYLYNMKQHGHLPSALPFENADA 882

QY 308 SQVDLNKBAQSIQTKIEQKITSNSTOWLRTAMEAFVAAQPKVKNMSTENFN-KGDHLQGG 366
DB 883 DPAELNHYSEIVQONIEKRISSETGNTDMLRTLMDHFTVNNPMWKNDSNVNFSGIFOQGG 942

QY 367 ALLYTHSDLTTPWANSYRLLNRTPTQ-QDGTKYFTFEGGEGGVFEFLISNDVDNSNPVQA 425
DB 943 FLKYNSDLTTPYANSYRLLGRPMINIKQOTYR-----GGEFLANDIDNSNPVQA 994

QY 426 EQLNLHLYLMWGDIVMGDKDANFDGVRVDAVDNVDNADLLQVYSNYFKDNYKVTSEANA 485
DB 995 EQLNLWLYLLNFCITITANNDQANFSDSVRVDAPDNIDADLMNIAQDYFNAAYGM-DSDAVS 1053

QY 486 LAHISILEAWSLNDNQYNDGTALSIDNSRLTSVLVTKQPGQORIDLNSLISSVNK 545
DB 1054 NKHINILEDWNHADPEYFNKIGNPQLTMDDTIK-NSLNHGLSDATNRWGLDAIVHQSL-A 1111

QY 546 BRANDTAYGDTIPTYSFVRAHDSVQTVIAKIVKEIDTNSDGYTFTLDQLDAFKIYNE 605
DB 1112 DRENNSTENVIPNYISFVRAHDNNSQDQIQAIRDV--TCKDYHTFTFEDEQKIDAIYIQ 1169

QY 606 DMAKVNTYTHYNI PAAYALLNSMESVPRVYVYGDLYTDGQVMAKSPYDAIAITMLOG 665
DB 1170 DQNSTVKKNLYNIPASYAILLTNKTDTIPRVYVYGDLYTDGQVMEHQTRYDYTLTLXKS 1229

QY 666 RIAYVSGQSSEVHKVNGNNQILSSVRYGODLMSADDTQGTDLRTSGLVTLVSNDPNLD 725
DB 1230 RVKYVAGGQSMQMSVGGNNILTSVRYGKGAMTATDT-GTDETRTGGIGVVVSNTPNLK 1288

QY 726 LG-GDSLTVNMGRHANOVRPLILGTQGVQSLKXSDTNIIVKYTDANGNLTFADD-- 782
DB 1289 LGVNDKVLVLMGAAHKNQQAARAVLTTDGVINITSQDQAP-VAMTDENGDLVLSHNLV 1347

QY 783 -----IKGYSYVDMSGYLAVVWVPVGAQDQGVVAAADTNQKADGKSLKTSAAlds 832
DB 1348 VNGKEADTAVQGYANPDVSGYLAVVWVPVGCASDNQDARTAPSTKXSGNSAYRTNAAFDs 1407
```

```

QY 833 QVIYEGFSNFQDFANNDADYTNKKIAENADFFKGLGITSFEMAPQVVSATDSFLDSIIQ 892
DB 1408 NIIFEAFSNFVYPTPKESERANVRIAQNAADFASLGFTSFEMAPQVNSKDRITFLDSTID 1467

QY 893 NGYAFSDRYDLAMSNNKYGSKDDLANALKALHANGIOAIADWVPDQIYQLPGEEVVTAk 952
DB 1468 NGYAFTRDYDLGMSPEPNKYGTDEDLRNALQALHAKGLQVMADWVPDQIYNLPKEVATV 1527

QY 953 RTNSYGNPTFDAYINNALLYATNTKSGSDYQAOYGAFIDELKAKYPDMFTNMISTGKP 1012
DB 1528 RVDDRGNVWKDAIINNLYVVNT-ICGGEYQKYGGAFLDKLQKLYPEIFTTKQVSTGVA 1586

QY 1013 IDPSTKIKOWEAKYENGNTNLGKAGYVLSDDATGKYFTV-NENGDFLPASTGD-QNAK 1070
DB 1587 IDPSQKITWESAQYFNGTNIHRGSGYVLKADG-QQYVNLGTTTKQFLPIQUTGEKKQGN 1645

QY 1071 TGFYYDGTGMAY-YSTSGNKAVNSFTYEG-GHYTYFDKGHMTVGSYKAEDGNDYVFLPN 1128
DB 1646 EGFVKNDGNYFYDLAGNWNVTFTEDSVGNWYFFDQDGKVENKH-----1692

QY 1129 GIQMRDAIYQDAGNSYYVGRGTGILYKGDNWTYFPVDPNNANKTVFRYFDANNVMAIGYN 1188
DB 1693 -----FVDVDS-----1698

QY 1189 MYGQ--TYFDENGFOAKQLLTDKGTGTHYFDENGAMAKNFVNVDGDDWYMDGNGNAV 1246
DB 1699 -YGEKGTFFLKNVGSFRGLVQTDNGTYFD-NYGMVNRQNTINAGAMIYTLDENGKLI 1756

QY 1247 KGQY 1250
DB 1757 KASY 1760
```

RESULT 13

```

US-09-604-957-4
; Sequence' 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAQUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-604-957-4
```

```

Query Match      19.9%; Score 1643; DB 2; Length 545;
Best Local Similarity 59.7%; Pred. No. 1.2e-91;
Matches 328; Conservative 73; Mismatches 124; Indels 24; Gaps 8;

QY 411 LLSNDVDSNPVQAEQLNLHLYLMWGDIVMGDKDANFDGVRVDAVDNVDNADLLQVYSN 470
DB 1 LLANDIDNSNPVQAEQLNLHLYLMWGDIVMGDKDANFDGVRVDAVDNVDNADLLQIASD 60

QY 471 YFKDNYKVTDSANALAHISILEAWSLNDNQYNDGTALSIDNSRLTSVLVTKQPG 530
DB 61 YLKAHYGVDKSEKNALNHLISILEAWSLNDNQYNDGTALSIDNSRLTSVLVTKQPG 530

QY 531 QRIDLSN-----LISESVNKERANDTAYGDTIPTYSFVRAHDSVQTVIAKIVE 580
DB 120 EK-DASNKNKEIRSGLEPVTNLSLN-NRSEAGKNSRMANYIFTRAHDSVQTVIAKIIA 177
```

581	Qy	KIDTNSDGYTFTLDOOLKQAFKIYNEDMAKUNTYTHYNI PAAYALLSNMESVPRVYGD	640
178	Db	QINPKTDELFTLDELKQAFKIYNEDMQAKKKYTQSNIPTAYALMSNKDSITRLRYD	237
641	Qy	LYTPDDGOYMAKSPYDAIAITWLOQRIAPVSGSQSEEVHKVNGNQ-----ILSSVR	692
238	Db	MYSDGQYMAKSPYDAIDITLLKARI KYAAGQDMKITVYEGDSKHMWDVYTGVLTVR	297
693	Qy	YGODLMSADDTQGTDLRSRTGLTVLTVSNDPNLDLG- GDSLTVNMGRAHANQYRPLILGT	751
298	Db	YGTGANEATD-QGSEATKTQGMAVITSNPSSLKLNQNDKVI VNMGAHKNOYRPLLLTT	356
752	Qy	KDGVOYILKDSDT-NIVKYTDANGMLTFTADDIKGYSTVDMSGYLAVVPVGAKQGDVR	810
357	Db	KDGLTSYTSDDAAAKSLYKRTNDKGBLVFPDASDIQGYLNPQVSGYLAVVPVGASDNQDVR	416
811	Qy	VAADTNQKADCKSLKTSALDSOVLTYEGFSNFQDFANNADADVTNNKIAENADFFKLIGIT	870
417	Db	VAASNKANATQVYESSALDSQLTYEGFSNFQDFVTNKDSYTNKKIAQNVLFRKSWGVT	476
871	Qy	SFENAPQVVSATDGSFGLDSIIIONGYAFSDRYDLAMSKNKYKGSKDDLANALKALHANGTQ	930
477	Db	SFENAPQVVSSEDSGFLDSIIIONGYAFEDRYDLAMSKNKYKGSQQQDMINAVKALHKSGLQ	536
931	Qy	AIADWVPDQ	939
537	Db	VIADWVPDO	545

```

RESULT 14
US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAQUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-995-749A-10

```

Qy	641	LYTDDGYMAKSPYYDAIATWLGRIAYVSGGSEEVHKVGNNO-----ILSSVR	692
Db	238	MYSDGQYMATKSPYYDAIDTLLKARIKAYAAAGQDMKITYYEGDKSHMDWDVTGVLTSVR	297
Qy	693	YGQDLMSADDTQGTDLSTSGLVTLVSNDFNLDLG-GDSLFTVMGRAHQAYRPLLGT	751
Db	298	YGTGANEAATD-QGSEATKTQGMAVITSNPPLKLNQNDKVIWNMGAHKNQYRPLLTT	356
Qy	752	KDGVCQSYLKQSDT-NIVKYTDANGLNFTADIIKG-YSTVDMSGVLAWVPVGAQGDGV	809
Db	357	KDGLTSTSDAAAKSLYRKNDKGELVFDASDIQGLYLNPOVSG-LAWVPVGAQDNQDV	415
Qy	810	RVAADTNQKADGSKLKTSAALDSQVIYEGFNFQDFANNADYTNKCTAENADFPKKLGI	869
Db	416	RVAASNKANATGVYESSALDSQIIYEGFNFQDFVTKDSQYTNKKIAQNVLFPKSGV	475
Qy	870	TSFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANALKALHANGI	929
Db	476	TSFEMAPQYVSSEDSGFLDSIIQNGYAFEDRYDLAMSKNNKYGQSQODMINAVKALHKSGI	535
Qy	930	QAIADWVPDQ	939
Db	536	QVIADWVPDQ	545

```

RESULT 15
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GREL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-11

```

Query Match	17.5%; Score 1442.5; DB 2; Length 522;
Best Local Similarity	56.1%; Pred. No. 1.7e-79;
Matches 298; Conservative	76; Mismatches 146; Indels 11; Gaps 7
QY	411 LLSNDVNSNPVQABQLNQLHYLMWGDITVMGDKANFDGVRVDADVNTNADLLQVYSN 470
Db	1 LLANVDNSNVVVEARQLNWLYLMMFGTITANDADANFDGIRVDADVNDVDRLDLQIARD 60
QY	471 YFKONYKVTDSSEANLAHISILEAWSLNDNQYNEDTNGTALSIDNSRLTSLAVLTQPG 530
Db	61 YPKLAYGVQNDATANQHLSSILEDSWSHNDPLYVTDQGSNQLTMDYDVHTQLIWSLTSSD 120
QY	531 QRIDLNLISSESVKSERANDTAYGDTIPTYSFVRAHDSSEVQTVAKIVKEKIDTNSDGYT 590
Db	121 IRGTMRQFV-DYYVMVDRSDSTENEAIPNYSFVRAHDSSEVQTVIAQVSDLYPQVENSIA 179
QY	591 FTLDQLKDAFKIYNEDMAVKNTKYTHYNI PAAYALLSNMESVPRVYGGDYTDGQYMA 650
Db	180 PTTEQLAAAFKVYNEDEKLADKKYTOYNMASAYAMLLTNKDTVPR-YGGDYTDGQYMA 238
QY	651 KKSPPYDAIATMLQGRIVYVSGQSEBEVHKVNGNQNILSSVRYGQDLWSADDDTGQDLSR 710

Db 239 TKSPYDAINTLLKARVQVAGQMSV-----DSNDVLTSVRYGKDAMTASDT-GTSETR 293
QY 711 TSGLVTVSNDPNLDL-GGDSLTVNMGRAHAQAYRPLILGTRKQGVQSYLKDSDTNI-VK 768
Db 294 TEGIGVIVSNNAEQLQLEDGHTVTLHMGAAHKKQAYRALLSTTADGLAYY--DTDENAPVA 351
QY 769 YTDANGNLFTTADDIKGYSTVDMSGYLAVWVPVGAKDQDVRVAADTNOKADGKSLKTS 828
Db 352 YTDANGDLIFTNESIYGVQNPQVSGYLAVWVPVGAAQDQDARTASDTTNTSDKVFSNA 411
QY 829 ALDSQVIYEGFSNFQDANNDADYTNKKTAEANADFEKLGITSEFEMAPOYVSATDGSPLD 888
Db 412 ALDSQVIYEGFSNFQAFATDSSEYTNVVAQNADQFKQMGVTSFQLAPQYRSSTDTSPLD 471
QY 889 SIIQNGYAFSDRYDLAMSNNKYGSKDDLANALKALHANGIOAIAADWVPDQ 939
Db 472 SIIQNGYAFTRDYDLGYGTPTKYGTADQLRDAIKALHASGIOAIAADWVPDQ 522

Search completed: February 11, 2006, 19:46:02
Job time : 42.0237 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 139.093 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-38
Perfect score: 8237
Sequence: 1 MEKKLHYKLHKVKKHWTTIA.....FVRIGDQWTFYFGNDGAATNL 1554

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_Main:*

1: /cgn2_6/prodata1/pubpaa/US07_PUBCOMB.pap:*

2: /cgn2_6/prodata1/pubpaa/US08_PUBCOMB.pap:*

3: /cgn2_6/prodata1/pubpaa/US09_PUBCOMB.pap:*

4: /cgn2_6/prodata1/pubpaa/US10A_PUBCOMB.pap:*

5: /cgn2_6/prodata1/pubpaa/US10B_PUBCOMB.pap:*

6: /cgn2_6/prodata1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8237	100.0	1554	US-10-383-930-38	Sequence 38, Appl
2	8237	100.0	1554	US-10-797-821-38	Sequence 38, Appl
3	3818	46.4	1518	US-10-383-930-40	Sequence 40, Appl
4	3818	46.4	1518	US-10-797-821-40	Sequence 40, Appl
5	3615	43.9	1430	US-09-740-274-6	Sequence 6, Appl
6	3615	43.9	1430	US-10-383-930-36	Sequence 36, Appl
7	3615	43.9	1430	US-10-797-821-36	Sequence 36, Appl
8	3502.5	42.5	1590	US-10-383-930-37	Sequence 37, Appl
9	3502.5	42.5	1590	US-10-797-821-37	Sequence 37, Appl
10	3434.5	41.7	1475	US-09-740-274-2	Sequence 2, Appl
11	3434.5	41.7	1475	US-10-383-930-34	Sequence 34, Appl
12	3434.5	41.7	1475	US-10-797-821-34	Sequence 34, Appl
13	3161.5	38.4	1375	US-09-740-274-4	Sequence 4, Appl
14	3161.5	38.4	1375	US-10-383-930-35	Sequence 35, Appl
15	3161.5	38.4	1375	US-10-797-821-35	Sequence 35, Appl
16	3080.5	37.4	1365	US-10-383-930-39	Sequence 39, Appl
17	3080.5	37.4	1365	US-10-797-821-39	Sequence 39, Appl
18	2917	35.4	1497	US-10-484-218-18	Sequence 18, Appl
19	2821	34.2	1595	US-10-484-218-20	Sequence 20, Appl
20	2422	29.4	2057	US-10-417-280A-2	Sequence 2, Appl
21	2285.5	27.7	1777	US-10-484-218-12	Sequence 12, Appl
22	2285.5	27.7	1781	US-09-995-749A-2	Sequence 2, Appl
23	2256	27.4	1006	US-10-484-218-22	Sequence 22, Appl
24	2199	26.7	1771	US-10-484-218-14	Sequence 14, Appl
25	1615	19.6	545	US-09-995-749A-10	Sequence 10, Appl
26	1442.5	17.5	522	US-09-995-749A-11	Sequence 11, Appl
27	1277	15.5	787	US-10-484-218-16	Sequence 16, Appl

28	1265	15.4	535	3	US-09-995-749A-13	Sequence 13, Appl
29	1172.5	14.2	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1158.5	14.1	525	5	US-10-484-218-23	Sequence 23, Appl
31	573	7.0	224	5	US-10-484-218-4	Sequence 4, Appl
32	556.5	6.8	223	5	US-10-484-218-10	Sequence 10, Appl
33	554.5	6.7	223	5	US-10-484-218-6	Sequence 6, Appl
34	545.5	6.6	2710	4	US-10-011-366-6	Sequence 6, Appl
35	545.5	6.6	2710	4	US-10-354-774-6	Sequence 6, Appl
36	545.5	6.6	2710	4	US-10-271-012-6	Sequence 6, Appl
37	545.5	6.6	2710	4	US-10-729-122-6	Sequence 6, Appl
38	545.5	6.6	2710	4	US-10-729-039-6	Sequence 6, Appl
39	545.5	6.6	2710	5	US-10-729-527-6	Sequence 6, Appl
40	545.5	6.6	2710	5	US-10-727-898-6	Sequence 6, Appl
41	545.5	6.6	2710	5	US-10-728-696-6	Sequence 6, Appl
42	545.5	6.6	2710	6	US-11-001-241-6	Sequence 6, Appl
43	510.5	6.2	221	5	US-10-484-218-8	Sequence 8, Appl
44	469	5.7	866	4	US-10-222-038-2	Sequence 2, Appl
45	469	5.7	866	5	US-10-987-508-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38

Query Match	100.0%	Score 8237;	DB 4;	Length 1554;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1554;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MEKKLHYKLHKVKKHWTTIAVASIGLVSLVGAGTVAEDKVANDTTAQATVGVDTGDDQA	60	
Db	1	MEKKLHYKLHKVKKHWTTIAVASIGLVSLVGAGTVAEDKVANDTTAQATVGVDTGDDQA	60	
QY	61	TTNDANTTTDTDTADQSANNTNQDQSGDSQNNQDQAKQDNTANTDRNQADNSQDNNQAT	120	
Db	61	TTNDANTTTDTDTADQSANNTNQDQSGDSQNNQDQAKQDNTANTDRNQADNSQDNNQAT	120	
QY	121	DOATSPATDGTSVQRDDAANVATAADQEQQTAPSEQESAAALSDNVKLDIGKYYVQAD	180	
Db	121	DOATSPATDGTSVQRDDAANVATAADQEQQTAPSEQESAAALSDNVKLDIGKYYVQAD	180	
QY	181	GSYKKNFALTNGQMLYFDSDFGALSSSTYSFSQGTTLNVDDPSSHNKAYDSTAKSFEL	240	
Db	181	GSYKKNFALTNGQMLYFDSDFGALSSSTYSFSQGTTLNVDDPSSHNKAYDSTAKSFEL	240	
QY	241	VNGYLTANSWTRPAGILRNGQTWEASNENDLRPLVMSWMPDKDQTOVAYVYNNKYLISANE	300	
Db	241	VNGYLTANSWTRPAGILRNGQTWEASNENDLRPLVMSWMPDKDQTOVAYVYNNKYLISANE	300	
QY	301	TEVNTNETSQVDLNKEAQSIQTQKIEQKITSQDNTSOMLRTAMEAFVAAQPKWNKSTENFNKG	360	
Db	301	TEVNTNETSQVDLNKEAQSIQTQKIEQKITSQDNTSOMLRTAMEAFVAAQPKWNKSTENFNKG	360	

QY 361 DHLQGGALLYTNSDLTPWANSYRLLNRTPTQODGTTKKTFTTEGEGGYEFLLSNDVDSN 420
DB 361 DHLQGGALLYTNSDLTPWANSYRLLNRTPTQODGTTKKTFTTEGEGGYEFLLSNDVDSN 420
QY 421 PVVQAEQLNQLHYLMNWGDI VMGDKDANFDGVRVDAVDNADLLQVYGNFYKDNKYKVD 480
DB 421 PVVQAEQLNQLHYLMNWGDI VMGDKDANFDGVRVDAVDNADLLQVYGNFYKDNKYKVD 480
QY 481 SEANALAHISILEAWSLNDQYNEEDTNGTALSIDNSSRLTSLAVLTQPKQORIDLNSLIS 540
DB 481 SEANALAHISILEAWSLNDQYNEEDTNGTALSIDNSSRLTSLAVLTQPKQORIDLNSLIS 540
QY 541 ESNVKERANDTAYGDTIPTYSFVRAHDSEVQTVIAKIVKEKIDTNSDGTFTFLDQLKDAF 600
DB 541 ESNVKERANDTAYGDTIPTYSFVRAHDSEVQTVIAKIVKEKIDTNSDGTFTFLDQLKDAF 600
QY 601 KIYNEDMAKNKTYTHYNIPAAVALLSNMESVPRVYVYGDLYTDDQYNAKKSPPYDATA 660
DB 601 KIYNEDMAKNKTYTHYNIPAAVALLSNMESVPRVYVYGDLYTDDQYNAKKSPPYDATA 660
QY 661 TMLQGRIAVYSGQSEVHKVNGNNOILSSVRYGQDLMSADDTQGTDLSTSLVTLVSN 720
DB 661 TMLQGRIAVYSGQSEVHKVNGNNOILSSVRYGQDLMSADDTQGTDLSTSLVTLVSN 720
QY 721 DPNLDJGGDSLTVNMGRANQAIRPLILGTGQVQSYLKSDSTNIVKYTDANGNLTFFTA 780
DB 721 DPNLDJGGDSLTVNMGRANQAIRPLILGTGQVQSYLKSDSTNIVKYTDANGNLTFFTA 780
QY 781 DDITKGYSTVDMSGYLAVVPVGAQGDQVRVAADTNQKADGKSLKTSAAALDSQVIEGFS 840
DB 781 DDITKGYSTVDMSGYLAVVPVGAQGDQVRVAADTNQKADGKSLKTSAAALDSQVIEGFS 840
QY 841 NFQDFANNADATYNNKIIAENADPFKKLIGITSFEMAPQYVSATDSGFLDSIIQNGYAFSDR 900
DB 841 NFQDFANNADATYNNKIIAENADPFKKLIGITSFEMAPQYVSATDSGFLDSIIQNGYAFSDR 900
QY 901 YDLAMSKNNKYGSKDDLALNALKALHANGIOAIADWVPDQIYQLPGBEVVTAKRTNSYGNP 960
DB 901 YDLAMSKNNKYGSKDDLALNALKALHANGIOAIADWVPDQIYQLPGBEVVTAKRTNSYGNP 960
QY 961 TPDAYINNALYATNTKSSGSDYQAOYGGAPLDELKAKYPDMFTVMNISTGKPIDPSKTK 1020
DB 961 TPDAYINNALYATNTKSSGSDYQAOYGGAPLDELKAKYPDMFTVMNISTGKPIDPSKTK 1020
QY 1021 QWEAKYFNGTNTVLGKAGYVLSDDATGKYFTVNENGDFLPASFTGQNAKTGFFYDGTGM 1080
DB 1021 QWEAKYFNGTNTVLGKAGYVLSDDATGKYFTVNENGDFLPASFTGQNAKTGFFYDGTGM 1080
QY 1081 AYTSTSGNAKAVNSFIYEGGHYYTFDKDGHMVTGQSYKABDNDYIFLPNGIQMRDAIYQDA 1140
DB 1081 AYTSTSGNAKAVNSFIYEGGHYYTFDKDGHMVTGQSYKABDNDYIFLPNGIQMRDAIYQDA 1140
QY 1141 QGNSYYTGRITGLYKGDNWPFFVDPNNANKTVPRYFDANNVMAIGYRNMVGTQYTFDENG 1200
DB 1141 QGNSYYTGRITGLYKGDNWPFFVDPNNANKTVPRYFDANNVMAIGYRNMVGTQYTFDENG 1200
QY 1201 FOAKGQLLTDDKTHYFDEDNAGAMAKNKFVNVGDDWYMDGNGNAVKGQYPVNNQILYFN 1260
DB 1201 FOAKGQLLTDDKTHYFDEDNAGAMAKNKFVNVGDDWYMDGNGNAVKGQYPVNNQILYFN 1260
QY 1261 PETGVQVKQGITDAQGRTSYYDANSALKSGSGFFTPNGSDWYIYAEVNGYVYKGFQKVAEN 1320
DB 1261 PETGVQVKQGITDAQGRTSYYDANSALKSGSGFFTPNGSDWYIYAEVNGYVYKGFQKVAEN 1320
QY 1321 QDQWYTFDQTTGQAKGAACVGRDLYFNPDSGVQVKGDFATDESNTSYFHGNDGKVV 1380
DB 1321 QDQWYTFDQTTGQAKGAACVGRDLYFNPDSGVQVKGDFATDESNTSYFHGNDGKVV 1380
QY 1381 GGFFTTGNNAWYADNNGNLVKGFQBEIDGKWYHFDVTCQAAGKALVNGQQLYFVDSG 1440
DB 1381 GGFFTTGNNAWYADNNGNLVKGFQBEIDGKWYHFDVTCQAAGKALVNGQQLYFVDSG 1440

QY 1441 IOVKGDFVTDGQNTSYDVNSGDKKXNGVFTTGDNAWYADGQGNLAKGRKSIDNQDLY 1500
DB 1441 IOVKGDFVTDGQNTSYDVNSGDKKXNGVFTTGDNAWYADGQGNLAKGRKSIDNQDLY 1500
QY 1501 FDPATGKQVKGQVLSIDGRNYYFDSGSGNMAKNRFRVIGDQWYIFGNDGAATNL 1554
DB 1501 FDPATGKQVKGQVLSIDGRNYYFDSGSGNMAKNRFRVIGDQWYIFGNDGAATNL 1554
RESULT 2
US-10-797-821-38
; Sequence 38, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J. .
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-38
Query Match 100.0%; Score 8237; DB 5; Length 1554;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1554; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKCLHYKLHKVKGHWVTIAVASIGLVSLVGAGTVAEDKVANDTTAQTATGVDTCQDQA 60
DB 1 MEKCLHYKLHKVKGHWVTIAVASIGLVSLVGAGTVAEDKVANDTTAQTATGVDTCQDQA 60
QY 61 TTNDANTNTTDTADQSAANTNODQAGSDQSNNOQAKODTANTDRNQADNSQTDNNQAT 120
DB 61 TTNDANTNTTDTADQSAANTNODQAGSDQSNNOQAKODTANTDRNQADNSQTDNNQAT 120
QY 121 DQATSPATDGTSVQRDAANVATAADQEGQTAPSEQEKSAALSLDNVKLIIDGKYYVQAD 180
DB 121 DQATSPATDGTSVQRDAANVATAADQEGQTAPSEQEKSAALSLDNVKLIIDGKYYVQAD 180
QY 181 GSYKKNFAITVNGQMLYFDSDTGALSSTSTYSFSGQTTNLVDDFSSHNKAYDSTAKSFEL 240
DB 181 GSYKKNFAITVNGQMLYFDSDTGALSSTSTYSFSGQTTNLVDDFSSHNKAYDSTAKSFEL 240
QY 241 VNGYLTANSWYRYPAGTLRNGQTVWEASNENDLRPVLMSWPPDKDTQVAYVNMKYLISANE 300
DB 241 VNGYLTANSWYRYPAGTLRNGQTVWEASNENDLRPVLMSWPPDKDTQVAYVNMKYLISANE 300
QY 301 TEVTNETSOVDLNKEAQSIQTKEQKITSNDNSTQWLTARAMEAFVAAQPKNMSTENFNKG 360
DB 301 TEVTNETSOVDLNKEAQSIQTKEQKITSNDNSTQWLTARAMEAFVAAQPKNMSTENFNKG 360
QY 361 DHLQGGALLYTNSDLTPWANSYRLLNRTPTQODGTTKKTFTTEGEGGYEFLLSNDVDSN 420
DB 361 DHLQGGALLYTNSDLTPWANSYRLLNRTPTQODGTTKKTFTTEGEGGYEFLLSNDVDSN 420
QY 421 PVVQAEQLNQLHYLMNWGDI VMGDKDANFDGVRVDAVDNADLLQVYGNFYKDNKYKVD 480

Db 421 PVQAEQNLHLYLWNGDI VMGDKDANFDCGRVDAVDNWNADLLQVSYNFKDNYKVT 480
Qy 481 SEANALAHISILEAWSLNDQYEDTNGTALSIDNSSRLTSLAVLTQPGORIDLSNLIS 540
Db 481 SEANALAHISILEAWSLNDQYEDTNGTALSIDNSSRLTSLAVLTQPGORIDLSNLIS 540
Qy 541 ESNVNERANDATYGDITPTYSFVRAHSEVQTVIAKIVKIDTNSDGYTFTLDQLKDAF 600
Db 541 ESNVNERANDATYGDITPTYSFVRAHSEVQTVIAKIVKIDTNSDGYTFTLDQLKDAF 600
Qy 601 KYNEDMAKNKVTYTHNIPAAVALLSNMESPRVYGYDLYDDGOYMAKSPYDAIA 660
Db 601 KYNEDMAKNKVTYTHNIPAAVALLSNMESPRVYGYDLYDDGOYMAKSPYDAIA 660
Qy 661 TMLQRIAYVSGGSEEHVKNQNLSSVRYQDLMSADDTQGTDLSTSGTLVLSN 720
Db 661 TMLQRIAYVSGGSEEHVKNQNLSSVRYQDLMSADDTQGTDLSTSGTLVLSN 720
Qy 721 DPNLDLGGDSLTVNMGRAHANOAYRPLILGTQGVQSYLKSDTNIIVKYTDANGNLTF 780
Db 721 DPNLDLGGDSLTVNMGRAHANOAYRPLILGTQGVQSYLKSDTNIIVKYTDANGNLTF 780
Qy 781 DDIGYSTVDMSCGLAVWVPVGAQGDVRAADTNQKADGKSLKTSAAALDSQVIYEGFS 840
Db 781 DDIGYSTVDMSCGLAVWVPVGAQGDVRAADTNQKADGKSLKTSAAALDSQVIYEGFS 840
Qy 841 NFQDFANNADATYTKKIAENADFPKLGITSPFMAPOYVSATDGSFLDSIIQNGYAFSDR 900
Db 841 NFQDFANNADATYTKKIAENADFPKLGITSPFMAPOYVSATDGSFLDSIIQNGYAFSDR 900
Qy 901 YDLAMSKNNKYGSKDDLALNALKALHANGIOAIADWVPDQIYQLPGBEVTAKRTNSYGNP 960
Db 901 YDLAMSKNNKYGSKDDLALNALKALHANGIOAIADWVPDQIYQLPGBEVTAKRTNSYGNP 960
Qy 961 TPDAYLNNALYATNTYKSSGSDYQAOYCGAFDELKAKYPMFTVMNISTGKPTDPSKTK 1020
Db 961 TPDAYLNNALYATNTYKSSGSDYQAOYCGAFDELKAKYPMFTVMNISTGKPTDPSKTK 1020
Qy 1021 QWEAKYFNGTNLGKAGYVLSDDATGKFTVNEGDFLPASFTGDNQAKTGYYPDGTGM 1080
Db 1021 QWEAKYFNGTNLGKAGYVLSDDATGKFTVNEGDFLPASFTGDNQAKTGYYPDGTGM 1080
Qy 1081 AYTSTSGNKAIVSYIEGHHYFYFDKGMVHTGYSYKAEDGNDYFPLNGIQMRDAIYQDA 1140
Db 1081 AYTSTSGNKAIVSYIEGHHYFYFDKGMVHTGYSYKAEDGNDYFPLNGIQMRDAIYQDA 1140
Qy 1141 QGNSYYGRTGILYKGDNYFPVDPNNANKTVPRYFDANNVMAIGYRNMVGTYYIPDENG 1200
Db 1141 QGNSYYGRTGILYKGDNYFPVDPNNANKTVPRYFDANNVMAIGYRNMVGTYYIPDENG 1200
Qy 1201 FOAKGQLLTDKGTHTYFDEBNGAMAKNFVNGDDWYMDGNGNAVKGOYPVNNQILYFN 1260
Db 1201 FOAKGQLLTDKGTHTYFDEBNGAMAKNFVNGDDWYMDGNGNAVKGOYPVNNQILYFN 1260
Qy 1261 PETGVQVKQFITDAQRTSYDANSALKSSGFFTPNGSDWYAEVGYVYKFKQVAEN 1320
Db 1261 PETGVQVKQFITDAQRTSYDANSALKSSGFFTPNGSDWYAEVGYVYKFKQVAEN 1320
Qy 1321 QDQWYFDQTTQKQAKGAKVGRDLYFNPDSGVQVKGFATDESNTSFYHGDNGDKVY 1380
Db 1321 QDQWYFDQTTQKQAKGAKVGRDLYFNPDSGVQVKGFATDESNTSFYHGDNGDKVY 1380
Qy 1381 GGFFTTGNNAWYADNNGNLVKGFBIDGKWHFDEVTQQAALVNGQQLYFDVDSG 1440
Db 1381 GGFFTTGNNAWYADNNGNLVKGFBIDGKWHFDEVTQQAALVNGQQLYFDVDSG 1440
Qy 1441 IQVGDFVTDGQNTSYDVNSGDKKVGFFTTGDNWYADQGNLAKGRKSIDNQDLY 1500
Db 1441 IQVGDFVTDGQNTSYDVNSGDKKVGFFTTGDNWYADQGNLAKGRKSIDNQDLY 1500
Qy 1501 FDPATGKQVKQLVSDGRNYFYDSSGSGNMAKNRFRIGDQWYIFGNDGAATNL 1554

Db 1501 FDPATGKQVKQLVSDGRNYFYDSSGSGNMAKNRFRIGDQWYIFGNDGAATNL 1554
RESULT 3
US-10-383-930-40
; Sequence 40, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-383-930-40

Query Match 46.4%; Score 3818; DB 4; Length 1518;
Best Local Similarity 50.1%; Pred. No. 2.4e-193;
Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;

Qy 1 MEKKLHYKLHKVKKHWTVIAVASIGLVSLVG-----AGTVSA---EDKVANDTTAQATVG 52
Db 1 MENKIHYKLHKVKKQWVTIAVASVALATVLGGLSVTTSSVSADETDQKTVTQNSGTTAS 60
Qy 53 VDTGQQAQNTDANTTTDTTADQSANNTQDQSGDQSNQQAQKQDQTANTORNOADNS 112
Db 61 LVTSPEATKEADKRTTKEADVLTPAKETNAVEATAT-TTNTQATAEAATATTADVA-VA 118
Qy 113 QTDNNQATQOATSPATDGTGSVQRDA---ANVATAADQEGOTAPSQEKSAALSLDNVKL 169
Db 119 AVPKAEAVTTDAPATTEKAEQPAIVKAEVVNTEVKAPEALKDSEVAALSLKNIKN 178
Qy 170 IDGKYYVQADGSYKKNFAITVNGQMLYFDSDTGALSSSTSYFSQGTNTLVDFDSSHKN 229
Db 179 IDGKYYVYVNDGSHKENFAITVNGQLLYFGKD-GALTSSTSYFTPTTNIVDGFSNNR 237
Qy 230 AYDSTAKSFELVNGLYTANSTRPAGILNGQVTEASNENDLRPVLMSWHPDKDTQVAVY 289
Db 238 AYDSSSEAPFELIDGYLTADSWYRPASIIKDGVTWQASTAEDFRPFLLMWPNVDTQVNYL 297
Qy 290 NYMNKYL SANETEVTNETSQVDLNKEAQSIQTKIEOKITSDNSTOWLRTAMEAFVAAQPK 349
Db 298 NYMSKVFNL-DAKYSSSTDQKQETLKVAADKIQIKIEQIOAEKSTQWLRETISAFVKTQPK 356
Qy 350 WNMSTENFNKG---DHLQGGALLYNTSLDTPWANSDYRLNLRPTQDGT--KKYFTEGG 404
Db 357 WNKETENYSKGGEDHLQGGALLYVNDSTPTWANSDYRLNLRNTATNTQGTIDKSIDDEQS 416
Qy 405 E-----GGYEFLLSNVDNSNPVQABQLNQLHYLNMWGDIVMGDKDANFDGVRVDAVDNV 460
Db 417 DPNHMGGFLLANDVDLSNPVQABQLNQLHYLNMWGSIVMGDKDANFDGVRVDAVDNV 476
Qy 461 NADLLQVTSNYPKDNKYKVTDSSEANALAHISILEANSNDNNOYEDNTNGTALSIDSRLT 520
Db 477 DADMLQLYNTYPREYYGVNKSSEANALAHISILEANSNDNHYNDKTDGAALAMENKORLA 536
Qy 521 SLAVLTK-----OPGQRIDLSNLISESVNKEBRAND-----T 551
Db 537 LLFSLAKPIKERTPAVSPLYNNFTFWTQDEKTDWLNKDGSKAYNEDGTVTKQSTIKYNE 596
Qy 552 AYGDITPTYSFVRAHSEVQTVIAKIVKIDTNSDGYTFTLDQLKDAFKIYNEDMAKNV 611
Db 597 KYGDAGSNVYFTRAHDNNVQDIIABIIKKEINPKSDGFTITDAEMKQAFIYNKMDMLSSD 656

```
QY 612 KYTHYNIIPAAVALLSNMESVPRVYDLYTDDGOYMAKSPYDYDAIATMLQRIAYVS 671
Db 657 KKYTLNIPAAVAVMLQNNMETITRVVYDLYTDDGHYMETKSPYDYDIIVNLKMSRIKYS 716
QY 672 GQSEEVH-----KVGNNQIILSSVRVYGDILMSADDTGDTLSRTSGIATLV 718
Db 717 GGOAQSYWLPTDGKWNDSVLEIRYNEVTSYRQDIWMTANDTEGSKYRSTSGGVTLV 776
QY 719 SNDPNLDLGDGS-LTYNMGRANHAQYRPLILGTGQVQSYLKDS- --TNIVKYTDANG 774
Db 777 ANPNKLNLDQSAKLVNEMGKIHANQYRALIVGTADGIKNFTSDADAIAAGYVKETDSNG 836
QY 775 NITFTADDIKGYSTVDMGYLAVVVPVGAQGDVRAADTNQADGK-SLKTSAAALDSQ 833
Db 837 VLTFGANDIKGYETDFMSGFVAVVVPVGAQSDNQDIRVAPSTEAKKEGELTLKATEYDSQ 896
QY 834 VIYEGSFNQDF--ANNADATYTKIAENADFKLIGITSFEMAPQVVSATDGSFLDSII 891
Db 897 LIYEGSFNQFTIPDGSDFSVYTNRKIAENVDLFKSGVTSFEMAPQVVSADDTGFLDSVI 956
QY 892 QNGYAFSDRYDLAMSNNKYSGKDDLANALKALHANGIOAIADWVPDQIYQLPGEVWTA 951
Db 957 QNGYAFADRYDLAMSNNKYSGKEDLRDALKALHKAGIOAIADWVPDQIYQLPGEVWTA 1016
QY 952 KRTNSYGNPTFDAYINNAYATNTKSSGSDYQAYGGAFLDELKAKYPMFTVMNISTGK 1011
Db 1017 TRTDGAGRKIADAIIDHLSLYVANSKSGKDYQAKYGGEGFLAELKAKYPMFKVMNISTGK 1076
QY 1012 PIDPSTKIKQWEAKYFNGTNVLKGAGYVLSDDATGKYTYVNGDFLPASFTGDQNAKT 1071
Db 1077 PIDDSVKLQKWAEBYFNGTNVLERGYYVLSDEATGKYFTYVKEGNFIFLQLTGKEKVI 1136
QY 1072 GFYDGTGMAYYSTGNKAVNSPIYEGGHYVYFDKGHMVTSYKAEADGNDYY-FLPNGI 1130
Db 1137 GFSSDKGITYFTGTSQAKSAFTNGTYTFDARGHMVTSYSPNGKDYVRFPLNGI 1196
QY 1131 QMRDAIYQDAQNSYVYRTGILYKGDNMVYFVDPNNANK-----TVFRYPDANNVMA 1183
Db 1197 MLSNAFYIDANGNTLYNSKGQMYKGG--YTKFDVSETDKDGESKVKAFRYFTNEGVA 1254
QY 1184 IGRNMYGOTYYPDENGFOAKGOLLTDDKTHYFDEDNAMAKNFVNGDDWYMDGNG 1243
Db 1255 KGVTVIDGTFQYFGEDGFOAKGLVTFKGKTYTFDAHTGNGIKOTWRNKGWYFPDANG 1314
QY 1244 NAVKGQYVNNQILYFNPTGVQVQGFITDAQRTSYDANSALKSGGFFTPNGSDWY 1303
Db 1315 VAATGAQVINGQKLYFN-EDGSQVKGVVKNADGTSYKKEGFGELVTFNEFFTIDGNVWY 1373
QY 1304 YAEVNGYVYKFKQVAENQDQWYFDQTTGQAKGAAGVGRDLYFNPDGQVKGDFATD 1363
Db 1374 YA-----GANGKTVTGAQVINGQHLVFNAD-GSQVKGWVKN 1409
QY 1364 ESGNTSFYHDNGDKVVGFFFTGNNAWYADNNGNLVKGFEIDGKWHFDEVTGQQAQ 1423
Db 1410 ADGTSYKYNASTGERLTNEFFFTGDNWYIYGANGSKVSGVEVKI-----1453
QY 1424 GAALVNGQQLYFDVDSGIOVKGDFVTDGQNTSYVDVNSGDKKVGNGFFFTGDNWYADG 1483
Db 1454 -----GDDTYFFAKDGKQVKGQTVSAGNGRISYYIGDGGKRAVSTWIEIQPGVYVYFDK 1507
QY 1484 QG 1485
Db 1508 NG 1509
```

RESULT 4

```
US-10-797-821-40
; Sequence 40, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
```

```
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
; US-10-797-821-40
```

```
Query Match 46.4%; Score 3818; DB 5; Length 1518;
Best Local Similarity 50.1%; Pred. No. 2.4e-193;
Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
```

```
QY 1 MEKHLHYKHLKVKHWIIVASIGLVSLVG-----AGTVSA---EDKVANDTTAQTAVG 52
Db 1 MENKHLHYKHLKVKHWIIVASVALATVILGSLVTSSTSSVADETDQKTVTQNSGTTAS 60
QY 53 VTDGQDQATNDANTNTDTDADQSANINQDQAGSDQSNQDQAKQDQTANTDRNQADNS 112
Db 61 LVTSPEATKEADKRTNKTEADVLTPAKETNAVETAT-TTNTQATAEAATTAITADVA-VA 118
QY 113 QTDNQAQTDQATSPATDGTSGVQRDA--ANVATAADQEGQTAPSEQESKAALSLDNVKL 169
Db 119 AVPNKEAVVTTDAPAVTTEKAEQPAIVKAEVYVNTVEVKAPEALDKDSEVEAALLSKNIKN 178
QY 170 IDGKYYVQADGSYKKNFAITVNGQWLYPDSDTGALSTSTSYFSOGTTNLNLDVDFSSHKN 229
Db 179 IDGKYYVNEDESHKENFAITVNGQLLYFGKD-GALTSSSTSYFTPTTINVDGFINNR 237
QY 230 AYDSTAKSPELVNGYLTANSWYRPAIGILNGOTWEASNENDLRPVLMSWMPDKDTQVAVY 289
Db 238 AYDSSEASPELIDGYLTADSWYRPASIIKDGVTWQASTAEDFRPLLMWMPNVDVTQVNYL 297
QY 290 NYMKYLSANETEVNTEISQVDLNKEAQSIQTKIEQKITSDNSTOWLRTAMEAFVAAQPK 349
Db 298 NYMSKVFNLD-DAKYSSSTDKQETLVKAAKDIIQIEQIOAEKSTOWLRETISAFVKTQPK 356
QY 350 WNMSTENFNKG--DHLQGGALLYTNISDLTPWANSDYRLNRTPTQDDGT--KKYFTEGG 404
Db 357 WNKETENYKGGEGDEHLQGGALLYNDSTRTPWANSDYRLNRTATNQTGTIDKSLDEQS 416
QY 405 E-----GGYEFLLSNVDNSNPVQAEQLNQLHLMNMGDI VMGDKDANFDGVRVDAVDNV 460
Db 417 DPNHMGFPDILLANDVDLSNPVQAEQLNQLHLMNMGSI VMGDKDANFDGVRVDAVDNV 476
QY 461 NADLLQVYSYFNKDYKVTDSEANALAHISILEANSINDNOYNEDTNGTALSIDSNSRLT 520
Db 477 DADMLQLYTNYPREYVGVNKSSEANALAHISVLEANSINDNHNKDTDGAALAMENKORLA 536
QY 521 SLAVLTK-----QPGQRIDLSNLSSESVNKERAND-----T 551
Db 537 LLFSLAKPIKERTPAVSPLYNNTFTTQDEKTDWLNKDGSKAYNEDGTGVKQSTIGKYNE 596
QY 552 AYGDPTIPTYSFVRAHDSYQTVIAKIVKEKIDPNSDGYFTTLDQLKDAFKYNEDMAKNV 611
Db 597 KYGDASGNVYFIRAHNNVQDIIAEIIKKEINPKSDGFTITDAEMKQAEFIYNKMDLSSD 656
QY 612 KYTHYNIIPAAVALLSNMESVPRVYDLYTDDGOYMAKSPYDYDAIATMLQRIAYVS 671
```


Db 649 SITRLYYGDMYSDGQYMATKSPYYDAIDTLKARIKYAAGQDMKITTYVEGDKSHMDW 708
QY 687 ---ILSSVRYGQDLMSADDTQGTDLRSRTSLGLVTLVNSDNLDLG-GDSILTVNMGRAHANQ 742
Db 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPFLKLNQNDKVINVMGAHKNQ 767
QY 743 AYRPLILGTGQGVQSVYKDSDT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPV 801
Db 768 EYRPLLLTTKDLGTLSTSDAAAKSLYRKNTDNGELVDFDASDIQGYLNPQVSGYLAVWVPV 827
QY 802 GAKDQDVRVAADTNQKADGKSLKTSAAALDSOVIVYEGFSNFQDEANNADYTNKKIAENA 861
Db 828 GASDNQDVRVAASNKANATGQVYESSALDSQLIYEGFSNFQDFTVKDSDYTNKKIAQNV 887
QY 862 DPFKGLGITSFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANAL 921
Db 888 QLFKSGWTSFEMAPQYVSSEDGSLDSIIQNGYAFEDRYDLAMSKNNKYGQQDMINAV 947
QY 922 KALHANGIQAIADWVPDQIYQLPGEVVTAKRTNSYGNFTFDAYINNALLYATNTKSSGSD 981
Db 948 KALKHSGIQVIAIDWVPDQIYNLPGEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGD 1007
QY 982 YQAOYGGAFDELKAKYPMDFVTNMI-STGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVL 1041
Db 1008 YQAKYGGAFSELAAKYPISFNRQTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVGYVL 1067
QY 1042 SDDATGKYFTVNEGDFLPASFTGDQNAKTGFYDGTGMAYTSTSGNKAIVNPFYBEG-GH 1100
Db 1068 KDNASKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFYTSQYQAKNSFVQDAKGN 1126
QY 1101 YYPFDKDHMTVGSYKAEDGNDYFPLPNGIQMRDALYQDAQNSYVYGTGILYKGNWY 1160
Db 1127 WYFDDNNGHMVYGLQOL-NGEVQYFLSNGVQLRESPLENADGSKNYFGLHGNRY-SNGYY 1184
QY 1161 PFVDPNNAKTYFRYEDANNWAIQYRNNYGTYYFEDENGFOAKGOLLTDDKG-THYRDE 1219
Db 1185 SF-----DNDKRWYFDSGVMAGVGLKTNGNTQYFDQDQYQVKGAWITGSDGKKRYFDD 1239
QY 1220 DNGAMAKNFVN-VGDDTYMDGNGNAVKGOYPVNNQIILYFNPETGVQVKGQFITDAQGR 1278
Db 1240 GSGNMAVRFANDKNGDWYLYNSDGLALVGVTNGKTYFG-QDGKQIKGKIITD-NGK 1297
QY 1279 TSYDANSALKSSGFTFNGSDWYEAENGYYVKGPKQVAENQDQWYFDTTQTKQAKGA 1338
Db 1298 LKYFLANSGLARNIPAT-----DSQNNWYFYG-SDGVAVTGS 1334
QY 1339 AKVDGRLYFNPDSGVQVKGFATDESNTSFVHGDNQKVVCGRPFTTGNNAWYADNNG 1398
Db 1335 QTIAGKKLYFASD-GKQVKGSEFT-YNGKVHYTHADSGGELQVNRFEADKDGKNWYLDNSG 1392
QY 1399 NLVKGQEI DGKWHYFDEVTGQQAALVNGQQLYFVDVDSGIVKQGFVTD 1450
Db 1393 EALTGSQRI-----NDQRVEF-TREGKQVKGDVAYD 1422

RESULT 6

US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIORITY FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36

; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-36
Query Match 43.9%; Score 3615; DB 4; Length 1430;
Best Local Similarity 49.9%; Pred. No. 1.2e-182; Indels 112; Gaps 32;
Matches 745; Conservative 217; Mismatches 418;
QY 1 MEKKLHYKLVKHKVHWVTIAVASIGLVSLVAGATVSAEDKVVANDTTAAQTAVGVDTGQ--- 57
Db 1 METKERYKMKHKVHWIVAVAS-GLIIL---GTTTLGSSVSAETEQQTSKVVVTKQSD 56
QY 58 DOATTNDANTNTTDTADQASANTNQDQASQSNNOQAKQDT-----ANTOR 106
Db 57 DKAASESQDTAPKT-----KQAQTEQTQAQS-QANVADTSTSTIKETPSQNTITQANSDD 111
QY 107 NOADNSQTDNNQATDOATSPATDGTSVQRRDAAVATA-ADQEGOTAPSEQEKSAAL-SL 164
Db 112 KTVVTNKSEBAQTSBERTKQSEEAQTTASSQALTQAKAELTKQRTAAQENKNPVDLAAI 171
QY 165 DNVKLIIDGKYVYVQADGSYKKNFAITVNGQMLYFDSDTGALSSTSYFSQGTTLNVDDF 224
Db 172 PNVKQIDGKYVYIGSDGQPKKNFALTNNKVLVYFKNTGALTDTTQYQPKQGLTKLNNDY 231
QY 225 SHNKAYDSTAKSFLVNGYLTANSWYRPAGILNRQGTWEASNENDLRPLVMSWPKDT 284
Db 232 TPNQIVNFENTSLTETIDNYVTADSWYRPKDILKNGKTWTASSESDLRPLMSWPKDT 291
QY 285 OYAVYNNKYLKANETETVNTNETSQVDLNKEAGSIQTKLEQKITSDNSQWLTAMEAFV 344
Db 292 QIAYLNNYMQGLGTGENYTDSSQESLNLAATQVQVKIETKISQTTQQTQWLRDIINSFV 351
QY 345 AAQPKWNNMSTE---NFNKGDLHQQGALLYTNSDLTPWANSYRLLNRTPTQODGTKKYFT 401
Db 352 KTOPNWSQTESDTSAGEKDLHQQGALLYSNSDKTAYANSYRLLNRTPTSTQTPKPYE 411
QY 402 EGEGEGYFLNSVDNPNVVAEQNLQHLMMWGDIVMGDKDANPFGVRVADVNVN 461
Db 412 DNSSGYDFLLANDIDNSNPVVAEQNLWHLMMYGSIVANDPEANPFGVRVADVNVN 471
QY 462 ADLLQVYNNYFKDNYKVTDDSEANALAHISLEAWSINDNQYNETNGTALSIDNSRLTS 521
Db 472 ADLLQIASLYLKAHYGVDKSEKNAINHLSILSAWSDNDPQYNKDTKGAQLPDKNKLRLSL 531
QY 522 LAVLTQPKQRIIDLNS-----LISESVNKERANDTAYGDTIPTYSFVRAHDSVQ 571
Db 532 LYALTR-PLEK-DASNKNEIRSGLEPVITNSLN-NRSABGKNSEMANIYIFIRAHDSVQ 588
QY 572 TVIAKIVKEIDTNSDGYFTFLDQKDAFKIYNEDMAKVNKTYTHYNIIPAAVALLLSNME 631
Db 589 TVIAKIIKAQINPKTDGLTFTLDELKQAFKIYNEDMRQAKKYTQSNIPTAYALMSLNDK 648
QY 632 SVPRVYVYGDLYTDDGQYMAKSPYYDATATMLQGRIVYVSGGQSEBHVKNVGNQ----- 686
Db 649 SITRLYYGDMYSDGQYMATKSPYYDAIDTLKARIKYAAGQDMKITTYVEGDKSHMDW 708
QY 687 ---ILSSVRYGQDLMSADDTQGTDLRSRTSLGLVTLVNSDNLDLG-GDSILTVNMGRAHANQ 742
Db 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPFLKLNQNDKVINVMGAHKNQ 767
QY 743 AYRPLILGTGQGVQSVYKDSDT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPV 801
Db 768 EYRPLLLTTKDLGTLSTSDAAAKSLYRKNTDNGELVDFDASDIQGYLNPQVSGYLAVWVPV 827
QY 802 GAKDQDVRVAADTNQKADGKSLKTSAAALDSOVIVYEGFSNFQDEANNADYTNKKIAENA 861
Db 828 GASDNQDVRVAASNKANATGQVYESSALDSQLIYEGFSNFQDFTVKDSDYTNKKIAQNV 887
QY 862 DPFKGLGITSFEMAPQYVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANAL 921
Db 888 QLFKSGWTSFEMAPQYVSSEDGSLDSIIQNGYAFEDRYDLAMSKNNKYGQQDMINAV 947

QY 922 KALHANGIOAIADWPDQIIVOLPGEVVTAKRTNSYGNPTFADVNNALYATNTKSGSD 981
DB 948 KALHKSQIADWPDQIIVLPGKEVVTATRVNDYGEYRKDSIKNTLYAANTKSGKD 1007
QY 982 YQAYGGAFLDELKAKYKYPDMFTVMISTGKPIDPSTKIKOWEAKYFNGTNVLGKAGYVL 1041
DB 1008 YQAYGGAFLSELAAKYPSIFNRTQISNGKKIDSEKITAWEAKYFNGTNILGRGVYVL 1067
QY 1042 SDDATGYFTVNGDPLPASTGQDNQAKTGFYDGTGMAYYSTSGNKAVNSFIYEG-GH 1100
DB 1068 KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFYSTSGYQAENSFVQDAKGN 1126
QY 1101 YYPFDKCHMVTGSKYKAEDGNDYFYPNGIQMDAIYQDAQCHSYNGRTGILYKGNWY 1160
DB 1127 WYFDFNNGHMYGLQQL-NGEVOYFLSNGVQLRESFLENADGSKNYPFGLGNRY-SNGYY 1184
QY 1161 PFVDPNNANKTVPYFDPANNVMAIGYRNMYGOTYFYPDENGFOAKGOLLTDDKG-THYFDE 1219
DB 1185 SF-----DNDSKRYFDPASGVMAVGLKTINGNTQYFDDGTYQVKGAWITGSDGKRYFDD 1239
QY 1220 DNGAMAKNFVN-VGDDWYTMGNGNAVKQYFVNNQILYFNPETGVQVKGQFITDAQGR 1278
DB 1240 GSGNMAVNRFPANDKNGDWYLYNSDGIALVGQVINGKTYFYG-QDGKQIKGKIITD-NGK 1297
QY 1279 TSYDANSALKSGSFTPNDSWYYAENGYYVYKGFQVAENQDQWYFPOTTGKQAKGA 1338
DB 1298 LKTYFLANSGBELARNIFAT-----DSQNNWYFYG-SDGVAVTGS 1334
QY 1339 AKVDGRDLYPNPSGVQVKGDFATDSGNTSFYHGONGDKVGGFFTTGNNANYADNG 1398
DB 1335 OTTAGKKLYPASD-GKQVKSFTV-YNGKRYHYHADSNGELQVNRFEADKDGNNWYLYDSNG 1392
QY 1399 NLVKGFEIDGKWHYFDEVTGQAKGAALVNGQQLYFVDVDSGIVQVKGDFVTD 1450
DB 1393 EALTGSQRI-----NDQRVFF-FREGKQVKGDVAYD 1422

RESULT 7

US-10-797-821-36
; Sequence 36, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-36

Query Match 43.9%; Score 3615; DB 5; Length 1430;
Best Local Similarity 49.9%; Pred. No. 1.2e-182;
Matches 745; Conservative 217; Mismatches 418; Indels 112; Gaps 32;

QY 1 MBKCLHYLHKVCHWVTIAVASIGLVSLVGAGTVAEDKAVANDTTAQTAVGVDTGQ---57

DB 1 METRRYKHKVCHWVTIAVAS-GLITL---GTTILGSSVSAETEQTSDKVVTKQSED 56
QY 58 DQATTTNDANTNTTDTTADQSANTNDQAGSDQSNQDQAKQDT-----ANTDR 106
DB 57 DKAASSQTDAPKT-----KQATQEQTAQS-QANVADTSTSTKETPSONITTOQANSDD 111
QY 107 NQADNSQTNNOATQATSPATDGTISVQRDAANVATA-ADQEQQTAPSQEAKSAAL-SL 164
DB 112 KTVNTKSEBAQTSEERTKQSEBAQTASSQALTKQAKAELTKQRTAAQENKPNVDLAAI 171
QY 165 DNVKLDGKYIYVQADGSKYKGNFAITVNGOMLYPDSDTGALSSTSTSYFSQGTNNLVDDF 224
DB 172 PNVKQIDGKYIYIGSDGQPKGNFALTVNNKVLFDKNTGALTDTSTQYQFQGLTKLNDY 231
QY 225 SSHKAYDSTAKSFELVNGVLTANSWVRPAGIILRNQGTWEASNENDLRPVLMSWPPDKOT 284
DB 232 TPNQIIVNFENTSLETIDNYVTADSWYRPNKIDILNKGKTWTASSEDRLPLMSWPPDKOT 291
QY 285 QVAYVNMKNKYL SANETEVNTNETSQVDLAKAEQSIQTKIEQKITSONSTQWLRTAMBAFV 344
DB 292 QIAYLNMNQOGLGTGENYTADSSQESLNLAAQTQVKIETKISQTOQTQWLRIINSFV 351
QY 345 AAOQKWNMSTE---NFNKGDLQGGALLYTNSDLTPWANSYDYLRLNRTPTQOQGTKKYFT 401
DB 352 KTOPNNSQTESDTSAGEXDHLQGGALLYSNSDKTAYANSYDYLRLNRTPTSTQTKPKYFE 411
QY 402 EGGEGGYEFLSNDVNSNPVQAEOLNOLHYLNMWGDIVMGDKDANFDGVRDADVNDVN 461
DB 412 DNSGGYDELLANDINDSNPVPVQAEOLNMLHYLNMYSIVANDPEANFDGVRDADVNDVN 471
QY 462 ADILQVYSPKONYKVTDSANALAHISILEAWSLNDNQYNDTNGTALSINDSSRLTS 521
DB 472 ADLQIATSDYLKHYGVDSKNAIINHLISILEAWSNDNDPOYNKDTKGAQLPIDNKLRLSL 531
QY 522 LAVLTQPGQRIDLSN-----LISESVNKERANDTAYGDTTPTYSFVRAHDSVQ 571
DB 532 LYALTR-PLEK-DASNKNBIRSGLEPVITNSLN-NRSAEGKNSERMANYIFIRAHDSVQ 588
QY 572 TVIATIKVKEIDTNSDGYTFTLDOLKDAKIKIYEDMAKNVKTTHYNIIPAAAYALLISNME 631
DB 589 TVIAKIIKAQINPKTDGLTFTLDELQAFKIYNEDMRQCKKYTQSNIPAYALMLSNKD 648
QY 632 SVPRVYGYDLYTDDGYMAKSPYDAIATMLQGRIAIYVSGQSEEVHKVNGNQ-----686
DB 649 SITRLYGYDMSDDGYMATKSPYDAIDTLKARIKAAAGGQDMKITTYEGGDKSHMDWD 708
QY 687 ---ILSSVRYGQDLSADDTQGTDLRSRTSLVTVLSNDPNLDLG-GDSLTVNMGRAHANQ 742
DB 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSNPSPSLKLNQNDKVIIVNMGAAHKQ 767
QY 743 AYRPLILGTGDKGVQSVYKDSDT-NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPV 801
DB 768 EYRPLLLTTKDGLTSYSDAAAKSLYRKNTNDKGELVFDASDIQGYLNPQVSGYLAVWVPV 827
QY 802 GAKDQDVRVAADTQKADGKSLKTSALDSQVITYEGFSNFODFANNDDYTNKKIAENA 861
DB 828 GASDNDQVRVAASNKANATGVYESSALDSQIIEGFSNFQDFTVKDSYTNKKIAQNV 887
QY 862 DFFKGLGITSFEMAPQVVSATDGSFLDSIIQNGYAFSDRYDLAMSNNKNGSKDDLANAL 921
DB 888 QLFKSGVTSFEMAPQVVSSESGSFLDSIIQNGYAFEDRYDLAMSNNKNGSKQDDMINAV 947
QY 922 KALHANGIOAIADWPDQIYQLPGBEVVTAKTNSYGNPTFADYINNALYATNTKSGSD 981
DB 948 KALHKSQIADWPDQIYNLPGEVVTATRVNDYGEYRKDSIKNTLYAANTKSGKD 1007
QY 982 YQAYGGAFLDELKAKYKYPDMFTVMISTGKPIDPSTKIKOWEAKYFNGTNVLGKAGYVL 1041
DB 1008 YQAYGGAFLSELAAKYPSIFNRTQISNGKKIDPSEKITAWEAKYFNGTNILGRGVYVL 1067
QY 1042 SDDATGYFTVNGDPLPASTGQDNQAKTGFYDGTGMAYYSTSGNKAVNSFIYEG-GH 1100

Db 1068 KDNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGN 1126
QY 1101 YYYFDKDGHWVTGSYKAEDGNDYYFLPNGIQMRDAILYQDAQNSYYTGRGLYKGDWY 1160
Db 1127 WYFYDNNGHMVYGLQOL-NGEVQYFLSNGVQLAESFLENADGSKNYFGLGNRY-SNGYY 1184
QY 1161 PFVDPNNAKTVPRYFDANNVMAIGYRNMVGYQYFYFDENGFOAKGOLLTDDKG-THYPDE 1219
Db 1185 SF-----DNDKRWYFDASGVMAVGLKTINGNTQYFDQGYQYKGAWITGSDGKKRYFDD 1239
QY 1220 DNGAMAKNFVN-VGDDWTYMDGNGNAVKQYFVNNQILYFNPETGVQVKGPITDAQGR 1278
Db 1240 GSGNAVNFANDKNGDWYLSNDSGLALGVQVITNGKYTFYG-QDCKQIKGKIITD-NGK 1297
QY 1279 TSYDANSALKSSGPTNGSDWYEAENGYYVYKGFQVAENQDQWYFQDTTQKQAKGA 1338
Db 1298 LKYFLANSGLARNIFAT-----DSQNNYYFG-SDGVAVTGS 1334
QY 1339 AKVDGRDLYFNPDSGVQVKGDFATDESNTSFYHGDNGDKVVGFFTTGNNAWYYADNG 1398
Db 1335 QTAGKGLYFASD-GKQVKSFTV-YNGKHYHADSGELQVNRFEADKXGNWYLDNSG 1392
QY 1399 NLVKGFQEIDGKWHFDEVTGQAKGAALVNGQOLYFVDSDGIVKGFVTD 1450
Db 1393 EALTGSQRI-----NDQVFF-TREGQVKGDVAYD 1422

RESULT 8
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match 42.5%; Score 3502.5; DB 4; Length 1590;
Best Local Similarity 45.4%; Pred. No. 1.3e-176;
Matches 749; Conservative 252; Mismatches 482; Indels 167; Gaps 40;
QY 1 MEKLLHYKLHKVKKHWTTAVASIG-LVSLVAGTYSAEKDVANDTTAQATVGVDTGQDQ 59
Db 1 MEKNVRFKHKVKKRWTTLSVASATMLASALGASVASADTDTASDDSNQAVW---TG-DQ 56
QY 60 ATTNDAA--NTNTTDTDTAQASNTQDQAGSDQSNQDQAKODTANTDRNQADNSQTDNN 117
Db 57 TTNNQATDQTSIAATATSEGSASTD---AATDQASAEQTQGTASTD--TAAQTNTNAN 111
QY 118 QATDQATSPATDG-TSVQRDAAANVATAAQDEGTAPSEBQEKSAALSLONVKLIDGKYYY 176
Db 112 EAKWVPTENENQGTDEMLAEAKNVATA---ESDSIPSDLAK-----MSNVKQVDGKYYY 163
QY 177 VQADGSKKNFAITVNGQMLYFSDTGALSSSTYSYFSQGTNNLVDD---FSSHNKAYDS 233
Db 164 YDQDGNVKNKNFAVSGDKIYYFD-ETGAYKDTSSKVDADKSSSAVSQNAATIFAANNRAYST 222
QY 234 TAKSFVLNGLYLVANSWYRPAGLRNGQTWEASNENDLRPVLMSWMPDKDTQVAYVYNNY 293
Db 223 SAKNFAVNYLTADSWYRPKSLKDGKTWESGKDDFRPLLMAWVDPDTETKENYVYNNY 282

QY 294 KYLSANETETVNTNETSQVDLNLKEAQSIQTKEOKITSDNSTOWLRTAMEAFVAAQPKWMS 353
Db 283 KVVGDKKT-YTAETSOADLTAAABLVOARIEOKISENTNKLWLRREAIASFVKTPQPMWNE 341
QY 354 TENFNKGDHLCQGALLYTN-SDLTPWANSDYELLNARTPTQODGTK-----KYFTEGEGGY 408
Db 342 SEK-PYDDHLQALLFPDQOTDLTPTQSNRYLLNARTPTNQTSGLDSRSTYFNPNDPLGGY 400
QY 409 EFLNSNDVNSNPVQAEQLNQLHYLMNWDIVMGDKDANFPGVRVDAVDNPNADLLQVY 468
Db 401 DELLANDVDNSNPVQAEQLNWLHYLLNFGSIYANDADANFDSIRVDAVDNVDADLLQIS 460
QY 469 SNYFKDNYKYTSEANALAHISILEAWSLANDQYNEDTNGTALSITDSSRLSLAVLTQK 528
Db 461 SDYLKAAAGIDKNNKANNHVSIVEAWSNDTFFYLHDDGDNILNMNDKFLSLMLSLAKP 520
QY 529 PQORIDLNLISSESYNKERANDTAYGDTIPTYSFVRAHDSVQTVIAKTVKEKIDTNSDG 588
Db 521 LDKRSGLNPLIHNSL-VDEVDREVEVTPSYSPARAHDSVQDIIRDIIKAEINPNSFG 579
QY 589 YFTTLDQLKAPKIYNEDMAKNKYTYTHYNIIPAAVALLSNMESVPRVYVYGLYTDGQY 648
Db 580 YSFTQBEIEQAFKIYNEDLKKTDKKYTHYNVPLSYLLLTNRKGSIPRVYVYGMFTDDGQY 639
QY 649 MAKKSPYDAIATMLQGRITAYUSGQSEBHVKNVNGNNOILSSRVRYQDLMASADDTQGTDL 708
Db 640 MANKTVNYDAIESLLKARKMKYSGGQAMQNYQI-NGGEILTYSRVYKGALQKSD-KGDAT 697
QY 709 SRTSGLVTLVSDPNLDLGGDSLTVNMGRAHQAQYRPLILGTKQGVQSYLKDSDTN--- 765
Db 698 TRTSGVGVNGQNPFSLDGKVVALNMGAAHQAQYRPLMVSTKQGVATYATDADASKAG 757
QY 766 IVKYTEANGNLFTTADDIKGYSTVDMSGYLAVMVPVGAQKGDQVRVAADTNQADKSKLK 825
Db 758 LVKRTDENGLYFLNDDLKGAVNPQVSGFLQVWVPVGAADDQDIRVAASDTASTDCKSLH 817
QY 826 TSAALDSQVIYEGFSPNFQFANNADADYTNKKTAEADAPFKKLGITSEFEMAPQVYSATDGS 885
Db 818 QDAAMDSTRVMEGFSNFQSFATKEBEYTNVAVANNVDFVSWGIGITDFEMAPQVYSSTDDG 877
QY 886 FLDSIIQNGYAFSDRYDLAMSNNKYGSKDDLALAKALHANGIOAIADWPDIQIYOLPG 945
Db 878 FLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWPDMQMTFFPK 937
QY 946 EEWVTAKTNSYGNPTFFDAYINNALLYATNTKSSGSDYQAYGGAFLDELKAKYPMFTVN 1005
Db 938 QEVVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDELKKEYPFLTKK 997
QY 1006 MISTGKPIDPSTKIKQWEAKYFNGTNNVLKGAGYVLSDDATGKYFTVNGDPLPASFTG 1065
Db 998 QISTGQAIQDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVASDITLFLPSSLIG 1057
QY 1066 DONAKTGFYDGTGMAYYST-SGNKAVNSFIYEGGHYVYFDKDGHWVTGSYKAEDGNDYY 1124
Db 1058 -KVVESGIRYDGGYIYNSSATGDQVQKASFIIEAGNLYYFGDKDGYNVTGA-QTINGANFY 1115
QY 1125 FLPNGIOMEDAIYODAQGNSYYTGRITLYK-----GDNWYFPVDPNNANKTVFRIYFDA 1178
Db 1116 FLENGTALRNTIYTDQAQGNSHYVANDGKRYENGYYQFGNDW-----RYFKD 1161
QY 1179 NNVMAIGYRNMVGYFYFDENGFOAKGOLL-TDDKTHYFDEDENGAMAKNFV-NVGDWW 1236
Db 1162 GN-MAVGLTVDGNYQYFDKGVQAKDKIIVTRDQKRVYFDQHNGNAANTFTIADKTGHW 1220
QY 1237 YYMDGNGNAVKGQYFVNNQILYFNPETGVQVKGFITDQAGRTSYDANSGLAKSSGFPT 1296
Db 1221 YYLKGDGVAVTGAQTGVKQKLYFE-ANGQVQKGDFTSDEGKLYFYFDVDSGDWMTDTTIE 1279
QY 1297 PNGSDGYAENGYYVYKGFQKVAENQDQWYFQDTTQKQAKGAQKVDGRDLRPNFDPGQV 1356
Db 1280 DRAGNWFYLGK-----DGAAVTGAQTIRGQKLYFKA-NGQOV 1315

QY 1357 KGPATDESNTSYHGDNGKXVGGFFTTGNNWYADNNG-----NLVKG----- 1403
Db 1316 KGDIVKGTGDKIRYIDAKSGEQVNKTVAADGKTYVIGNDGAVDPVVKGTQFKDASG 1375
QY 1404 ---PQEDGK-----NY---HEDEV---TCQAKGAALVNGQOLYFDVDSGTQVKGDFV 1448
Db 1376 ALRPNLKGQVLVTGSGWYETANHDWVYIQSGKALTGEQTNGOHLYPKED-GHQVKGQVL 1434
QY 1449 TDGQGNSTYYDVNSGDKKXVAGFFTTGDNWY----- 1480
Db 1435 TGTGDKVRYDANSQDQAFNKSVTVNGKTYVFGNDGTAQTAGNPKGTQFKDGSDFRYSYM 1494
QY 1481 -----ADGQ-----GNLAKGRKSIDNQLYFDPATGKQVKGQVLV-SIDGRN 1520
Db 1495 EGQLVTGSGWYENAGQGLVYXNGKVLTLGLTVGSQRVYFD-ENGIQAKGKAVRTSDGI 1553
QY 1521 YPDSGSGNMAKRFVRIGDQWLYFGNDGA 1550
Db 1554 RYFDNSGSMITNQWKFYQYQYFFGNDGA 1583

RESULT 9
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797, 821
; PRIOR FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match 42.5%; Score 3502.5; DB 5; Length 1590;
Best Local Similarity 45.4%; Pred. No. 1.3e-176;
Matches 749; Conservative 252; Mismatches 482; Indels 167; Gaps 40;

QY 1 MEKHLKHLKVKHVVITAVASIG-LVSLVAGAGTVSAEDKVANDTTAQATVGVDTGQDQ 59
Db 1 MEKNVRFKHKVKRVRWTVLSVASATMLASALGASVASADTTDASDSSNQAVV---TG-DQ 56

QY 60 ATTNDAA--NTNTTDDTADQSANTNQDQAGSDQSNQDQAKQDPTANTDRNQADNSQTDNN 117
Db 57 TTNNQATDQTSIAATATSEQASTD--AATDQASAEQQTGTTASTD--TAAQTNTNAN 111

QY 118 QATDQATSPATDG-TSVQRRDAANVATAADQEQATPSBOEKAALSLONVKLIDGKYYY 176
Db 112 EAKVVPTEENQGTDEMLAEAKNVATA---ESDSIPSDLAK-----MSNVKQVDGKYYY 163

QY 177 VQADGSYKKNFATVNGQMLYFDSDTGALSSTYSFSQGTNNLVDD---FSSHNKAYDS 233
Db 164 YDQDGNVKNKFNFAVSGDKIYYFD-ETGAYKDTSKVDADKSSSAVSQNAITFAANNRAYST 222

QY 234 TAKSFELVNGYLTANSWYRPAGLRNGQTWEASNENDLRPLVMSWMPDKDTQVAVYVYWN 293

Db 223 SAKNFEAVDNYLTADSWYRPKSLKDGKTWTEGSKODFRPLLMAWVPDTETKRYVYVNM 282
QY 294 KYLSANETEVTNETSQVDLNKEAQSQTKEQKITSNDSTOWLRTAMEAFAVAQPKNMS 353
Db 283 KVGVIDKT-YTAETSQADLTAAAEVQARIEQKITSNTNKKWLEAISAFAVKTPQOWNGE 341
QY 354 TENFNKGDLHQGALLYTN-SDLTPWANSDYRLNRPPTQDDGTK-----KYFTEGGGGY 408
Db 342 SEK-PYDDHLQNCALLFDNQDTLTPDQSNYRLNRPPTQDGLSDSRFTYVNPNDPLGGY 400
QY 409 EFLSDVNDNSNPVQAEQLNQLHLYLMNMGDIWVGDKDANFDPVVRVDAVDNADLLQVY 468
Db 401 DFLANDVDNSNPVQAEQLNQLHLYLMNMGDIWVGDKDANFDPVVRVDAVDNADLLQVY 460
QY 469 SNYFKDNYKYVDSEANALAHISILEAWSLNDNOYEDTNGTALSIDNSSRLTSLAVLTQ 528
Db 461 SDYLKAAAYGLDKNNKANNHVSIVEAWSNDTPLYLHDDGDNLMNKNKFLSLMLWSLAKP 520
QY 529 PGQRIDLSNLISSEVANKERANDTAYGDTIPTYSFVRAHDSVQTVIAKIVKEKIDTNSDG 588
Db 521 LDKRSGLNPLJHNSL-VDREVDREVEVPSYSFAHDSVQDIIRDIIKAEINPNPSFG 579
QY 589 YFTFLDQLKDAFKIYNEDMAKNKTYTHYNI PAAYALLLSNMESVPRVYVYDLYTDDQY 648
Db 580 YSFTQEEIEQAFKIYNEDLKTKTYTHYVPLSYTLTLTKGSI PRVYVYDLYTDDQY 639
QY 649 MAKSPYVDAIATMLQGRATAYVSGGSEEVHVKVNGNNOILSSVRYGQDLMSADDTQGTDL 708
Db 640 MANKTYNYDAIESSLKARMKYVSGQAMQNYQI-GNGEILTSVRYGKALUKQSD-KGDAT 697
QY 709 SRTSGLVTLVSDNPDLGDSLTVMNGRAHQAAYRPLILGTGKQVQSLYKQSDTN--- 765
Db 698 TRISGVVVMGQPNFSLDGKVALNMGAAHQAAYRPLILGTGKQVQSLYKQSDTN--- 757
QY 766 IVKYTDANGLTFTADDIKGYSTVDMGSLAVVWVPGAKDQDVRVAADTNQKADGKSLK 825
Db 758 LVKRTDENGVLFLNDDLKGVANPQVSGFLQVWVPGVGAADDQDIRVAASDTASTDGKSLH 817
QY 826 TSAALDSQVYEGFSNFPDANNADYTNKKAENADFFKGLGITSFEMAPQVVSATDGS 885
Db 818 QDAAMDSRVFMFGFSNFPQSFATKBEETVNVVIANNVDFKPSWGITDFEMAPQVVSSTGQ 877
QY 886 FLDSIIQNGVAFSDRYDLAMSNNKYGSKDDLALNALKALHANGIOAIADWVDPQIYQLPG 945
Db 878 FLDSVIONGYAFTDRYDLGSKANKYGTADQLVKALKALHAKGLKVMADWVPMQMTFFPK 937
QY 946 EEVVTAKRTNSYGNPTFFDAYINNALYATNTKSGSDYQAYGGAFDELKAKYPMFTVN 1005
Db 938 QEVVTVTRTDKFGKPTAGSQINHSLVYTDTKSGDDYQAKYGGAFDELKAKYPELFTTK 997
QY 1006 MISTGKPIDPSTKIKOWEAKYFNGTVNLGKAGYVLSDDATGKYFTVNGNDFLPASFTG 1065
Db 998 QISTGQADPSPVKIKQWSAKYFNGSNILGADYVLSQVSNKYFNVASDTLELPSLLG 1057
QY 1066 DONAKTGFYDGTGMAYYST-SGNKAVNSFIYEGGHYVFDKDHVMTGYSYKAEADNDY 1124
Db 1058 -KVVESGIRYDKGYIYNSSATGQDVQKASFIETAGNLYYFGKDYMTGA-QTINGANYF 1115
QY 1125 FLPNGIQMRDAIYQDAQGNSSYYGRTGILYK-----GDNWYFVDPNNANKTVPRYFDA 1178
Db 1116 FLENGTALRNTIYTDAGNSHYVANDGKRYNGYQFGNDW-----RYFKD 1161
QY 1179 NNVMAGYRNWYQTYVFDENGFAQKQLL-TDDKTHYFDEEDNGAMAKNFV-NVGGDW 1236
Db 1162 GN-MAVGLTVBGNVQYFDKGVQAKDKIIVTRDGVRYFDQHNAGNAATNTFIADKTGH 1220
QY 1237 YVMDGNGNAVKGQYVNNQILYFNPETGVQVKGQRTTDAQGRTSYVDANSALKSSGFFT 1296
Db 1221 YYLKDKGVAVTGAQTVGKQKLYFE-ANGQOVKGDFVTSDEGKLYFYDVSQDMMWTTFFIE 1279
QY 1297 PNGSDWYYAENGYYVYKGFQKVAENQDQWYFDQTTGKQAKAAKVDGRDLYFNPDGQVQ 1356

```
Db 1280 DRAGNWFYLGK-----DGAAVTGAQTIRGQKLYFKA-NGQV 1315
QY 1357 KGFAPDESGNTSFYHGDNGDKVVGFFTTGNNAWYADNG-----NLVKG-----1403
Db 1316 KGDIVRGTDKIRYDAKSGEQVFNKTVKAADGKTVIIGNDGVAVDPVSVKGTQFKDASG 1375
QY 1404 ---FQEIDGK-----WY---HFDEV---TGQAKGAALVNGOOLYFDVDSGIVKQKDFV 1448
Db 1376 ALRFYMLKQOLVTGSGWYETANHDWYIOSGKALTGEQTINGOHLFYKED-GHGVKGQJLV 1434
QY 1449 TDQGNNTSYVDNSGDKVNGFFTTGDNWY-----1480
Db 1435 TGTDGKRVYDANGSGDQAFNKSVTVNGKTYFNGDGTAGTGNPKGQTFKDGSDIRFYSM 1494
QY 1481 -----ADGQ-----GNLAKGRKSIDNQDLYFDPATGKQVKGQJLV-SIDGRN 1520
Db 1495 EGQLVTSGSWYENAOQWLYVXNGKVLTLGLQTVGSQRVYFD-ENGIOAKGKAVRTSDGKI 1553
QY 1521 YYPDSGSGNMAKRFVRIGDOWLYFGNDGA 1550
Db 1554 RYFDENSGSMITNQWKFVYGGYYFGNDGA 1583

RESULT 10
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US2002031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740, 274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 41.7%; Score 3434.5; DB 3; Length 1475;
Best Local Similarity 46.6%; Pred. No. 4.6e-173;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;

QY 1 MEKKLHVKLHVKKHVVTVAVAS-IGLVSLVGAGTVSAEDKVANDTTAQTAVGVDGQDQ 59
Db 1 MDKKVRYKURKKVKKVWTVSVSAVMTLTLSGLVKAD---SNESKQ-----46
QY 60 ATTNDANTNTTDTADQSAN-----TNQDQAGSDQSNQDQAKQDTAN-----103
Db 47 -ISNDSNTSVV---TANESNVITEATSKQEAASSQTNHTVTTSSSSTSVNPKVSVNP 102
QY 104 -TDRNQADNSQTDNNQATQATSPATDGTSVQRDDAANVATAADQGGQTAPSEQKSAAL 162
Db 103 YTVGETASNGEKLQNTT-----TVDKTS-----EAAANNISKQTTTEADTDVIDDSNAANL 153
QY 163 ----SLDNVKLIDGKYVYVQADGSYKKNFAITVNGOMLYFDSPTGALSSTSTYSFSQG-T 217
```

[illegible]

Qy	338	TAMEAFVAAQAKWNMSTEN-FNKGDLHOGGALLYTN-SDLTPWANSYRLNRRPTQDG	399
Db	332	QTISAFVKTQSAWNSDSEKPPD--DHLQNGAVLDNEGKLTPTYANSNRIILNRTPTNQTG	389
Qy	396	TK--KYFTEGEGGGEYFLLSNDVNSNPVQAEQLNQLHLMLNMGDIYVMGDKDANFCGVR	453
Db	390	KDPRYTADNTIGGYEFLANDVNSNPVQAEQLNQLHLMLNFGNIYANDPDANFDSIR	449
Qy	454	VDADVNNADLLQVYSNTYFKONYKVYTDSEANALAHISILEAWSLNDNOYNEDTNGTALS	513
Db	450	VDADVNDADLLQIAGDYLKAAGHKNDKAANDHLSILEAWSNDNTPYLDHDDGDNMINM	509
Qy	514	DNSRLTSLAVLITKOPQORIDLSNLSSESVKERANDTAYGDTTPYTSFVRAHSEVQTV	573
Db	510	DNKRLSLLFSLAKPLNORSQWMLTNSL-VNRTDDNAETAAPVSPSFIRAHDSVEVDL	568
Qy	574	IAKIYKEIKDNTSDGTYFTFDLQADAFKIYNEDMAKVNKTYTHYNIIPAAAYALLSNMESV	633
Db	569	IADIIKABINPVGVYSTMEIKAFIYNKDLATEKKYTHNTALSUALLTNKSIV	628
Qy	634	PRVYGDLYTDDGYMAKSPYDAIATMLQRIAYYSGGQSEEHVKHNGNNQILSSVRY	693
Db	629	PRVYGDMTDDGYMAKHTINYEALTELLKARIKYVSGGQAMRQOV-GNSEIITSVRY	687
Qy	694	QDLMASDDTGTDLRTSGLVTLVSDNPDL-L-GDGLTVMNGRAHANQAYRPLILGTK	752
Db	698	KGKALKATDT-GDRTRTTSRGVAVIEGNNPSLRKASDRVVVMGAAHKNOAYRPLLLTTD	746
Qy	753	DGVOSYLKQSD-TNIVKYTDANGNLTFTADDIKGYSTVDSMGSYLAVWVPICAK--DGQDV	809
Db	747	NGIKAYHSDQBAAGLVRYTDRGELIFTAADIKYANPQVSGYLGWVPVGAALIKMPAL	806
Qy	810	RVAADTQKADGKSLKTSAAALDSQVITYEGFSNFQDFANNADATYTKKIAENADPFCKLGI	869
Db	807	RLARPHQOMA---SVHQNALDSRVMEGFSNFQAFATKKEEYTNVIAKNDVFAEWGV	863
Qy	870	TSPFMAQYVSATGCSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANKALHANGI	929
Db	864	TDFEMAQYVSSTDGCSFLDSVIQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSKI	923
Qy	930	QAIADWPDQIYQLPGEENVTAKTNSYGNPTPDAYINNALYATNTXSSGSDYQAOYVGA	989
Db	924	KVMADWPDQYAPPEKEENVTAIRVDKYGTFVAGSQIKNTLYVVDGSSGKQOQAKYVGA	983
Qy	990	FLDELKAKYPMFTVMNMISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSLSDATGY	1049
Db	984	FLELQAKYPELFARKQISTGVPMDPVSKIKQWSAKYFNGTNILCRGAGYVLKQOQANTY	1043
Qy	1050	FTVNEGD--FLPASFTGDQNAKTFYYDGTGMAYYSTSGNKAVNSPIYEGGHYYFDDK	1107
Db	1044	FNISDNKEINFLPKTLL-NQDSQVGSFYDGKGYVYVYSTGSVQAKNTFISEGDKWYFDNN	1102
Qy	1108	GHMVTGSYKAEDGNDYVFLPNGIQMRDAIYQDAQGNISYYGRTGILYKGNWNYPFVDPNN	1157
Db	1103	GYMVTGA-QSINGVNYVFLSNGQLRDAILKNEIDGTAYYVGNDRRYE-NGYQFM----	1156
Qy	1168	ANKTVPRYFDANNVMAIGYRMVYGOTVYFDENGFOAKGOLLTDDKG--THYFDEEDGAWAK	1226
Db	1157	--SCVWRHFV-NGEMSGLTVIDGOVQYFDPMGFOAKGKFTTTADGKIRIFDQSGNNYR	1213
Qy	1227	NKFV-NYGDWYIMDGNNGNAVKGQYVNNQILYFNPETGVQVKGQFITDAQRTSYVDAN	1285
Db	1214	NRFTENEEGKWLVLGEDGAAVTSQTINGQHLVFR-ANGVOVKGEFVTDHHRISYYDGN	1272
Qy	1286	SGALKSGGFTPNGSWDYYAENGYYVYKGFQVQVAENQDQWYTFDQTTQKQAKGAUKDGRD	1345
Db	1273	SGDQIRNRF-----VRNAQOQWYFVD-NNGYAVTGARTINGQL	1309
Qy	1346	LYFNPDGSOVKGDPEATDESGNTSFYHGDNGDKVGVGFFTTGNNAWYADNNGNLVKGFQ	1405
Db	1310	LYFRA-NGVOVKGEFVTDYGRISYDNGNSGDQIRNRFVRNAQOQWYFVNNGYAV----	1364
Qy	1406	EIDCKWTHFDEVTCQAKGAALVNGQOLYFEDVDSGIVQKGDFTVDGOGNTSYVDVNSGDK	1465

Db 1365 -----TGARTINGQHLYFRA-NGVQVKGFEVTDHGRISYDNGSGDQ 1406
QY 1466 KVGPFPTTGNWYADGOGNLAGKRSIDNODLYDPATGKQVKGQLYSID-GRNYEYD 1524
Db 1407 IRRFVRNAQGWYFDNNGYAVTGARTINGQHLYF-RANGVQVKGFEVTDYGRISYYD 1465
QY 1525 SGSGNMAK 1532
Db 1466 ANSGERVR 1473
RESULT 12
US-10-797-821-34
; Sequence 34, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-34

Query Match 41.7%; Score 3434.5; DB 5; Length 1475;
Best Local Similarity 46.6%; Pred. No. 4.6e-173;
Matches 731; Conservative 248; Mismatches 458; Indels 131; Gaps 39;
QY 1 MEKLLHYKLHKVKKHVVITAVAS-IGLVSLVGAGTVSAEDKVVANDTTAQTAVGVDTGQDQ 59
Db 1 MDKKVRYKLRVKRWVTVSVASAVMTLTTLGGGLVKAD---SNESKSQ----- 46
QY 60 ATTNDANTNTDTDADQSAN-----TNODQAGSDQSNQDQAKQDPTAN----- 103
Db 47 -ISNDSNTSVV---TANESNVITEATSKQEAASSQTNHVTSSSSTSVNPKVSVNP 102
QY 104 -TDRNQADSQTDNNQATQATSPATDGTGTSVQRDRDAANVATADQBGQATPASEQKSAAL 162
Db 103 YTVGETASNGEKLQNTT-----TVDDKTS---EAAANNISKQTTEADTDVIDDSNAANL 153
QY 163 -----SLDNVKLIDGKYVYVQADGSYKKNFAITVNGQMLYFSDSDTALSSTSTVSFSQG-T 217
Db 154 QILEKLPNVEIDGKYVYVYDNNQKVRNTFTLIADGKILHFD-ETGAYTDTSIDTVNKOIV 212
QY 218 TNLVDDFSHKNKAYDSTAKSFELVNGYLTANSWYRPAGILRNGQTWAEASNENDLRPVLMS 277
Db 213 TTRSNNLYKKYNQVYDRSAQSFEHVDHYLTAEWSYRPKYILKDGKWTWTQSTEXDFFLLMT 272
QY 278 WPPDKDTQVAYVNNMKYLSANETEVNTSVDLNLKEAQSQTQKTEQKITSNQTQWL 337
Db 273 WPPDQETQRYVNNMAQLGINKT-YDDTSNQLNLNIAAAATQAKTEAKITTLKNTDWLR 331
QY 338 TAMEAFVAAQPKWNMSTEN-FNKGDLHQGGALLYTN-SDLTPWANSDYRLNARTPTQDQ 395

Db 332 QTISAFVKTQSAWNSDSEKPPD--DHLQNGAVLYDNEGKLT.PYANSNYRILNRTPTNQTG 389
QY 336 TK--KYFTGEGGGYEFLLSNDVNSNPVQABQLNQLHYLANWGDIWVGDKDAHPDQVGR 453
Db 390 KKDPRTADNTTGGYEFLLANDVNSNPVQABQLNQLHYLANWFGNIYANDPDANFDSIR 449
QY 454 VDADVNVADLLOVYSNYFKDNYKYTDSEANALAHISILEAWSLNDNQNYEDTNGTALGI 513
Db 450 VDADVNVADLLOIAGDYLLKAAKGTHKNDKKAANDHLSILEAWSNDNTPYLHDGDNMIMN 509
QY 514 DNSSRLLTSIAVLTKOPGORIDLSNLISBSVKNKERANDTAYGDTIPTYSFVRAHDSVQTV 573
Db 510 DNKLRLSLFLSLAKPLNQSRSGMNPJITNSL-VNRTDDNAETAAPSYSIFRAHDSVQDL 568
QY 574 IAKIVKEKIDTNSDGYTFTLDOLKDAFKIYNEDMAKVNKTYTHYINIPAYALLLSNWSV 633
Db 569 IADIIKAEINPNVVGYSFTMBEIKKAFETYNKDLATEKKYTHYNTALSYALLLTNKSIV 628
QY 634 PRVYVYDLYTDDGQYMAKSPYDIAIATMLQGRIAVVGSGGSEEVHVKVNGNNOISSLVRY 693
Db 629 PRVYVYDLYTDDGQYMAKHTINYEAITLLKARIKTVSGGQAMRNOQV-GNSEIITSVRI 687
QY 694 GODLMSADDTQGTDLRSGLVTLVSNPNLNL-GGDSLTVNMGRANQAYRPLITLTK 752
Db 688 GKGALKATDT-GDRTTTSVAVIEGNNPSLRLKASDRVVVNMGAHKNQAYRPLLLTTD 746
QY 753 DGVSYLKDSQ-TNIVKYTDANGNLTFTADDIKGYSTVDMSGYLAVVVPVGAQ--DGQDV 809
Db 747 NGIKAYHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGYLVGVVPGAALIKMPAL 806
QY 810 RVAADTNQKADGKSLKTSNAALDSOVIYEGSFNFQDPANNDADYTNKKTAENADFCKLGI 869
Db 807 RLARPHQQA---SVHQNAALDSRVNMFEGSFQAFATKKEBYTNVTAKNVDFKFAEWSV 863
QY 870 TSFEMAPQVVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKTKSGKDDLANALKALHANGI 929
Db 864 TDFENAPQVVSSTDGSFLDSVILQNGYAFTRDYDLGISKPNKYGTADDLVKAIKALHSKGI 923
QY 930 QAIADWVPDQIYQVLPGEVVTAKRTNSYGNPTFFDAYINNALLYATNTKSSGSDYQAOYGA 989
Db 924 KVMADWVPDQVYAPPEKEVVTATRVDKYGTVPVAGSQIKNTLYVVDGKSSGKQQAQYGA 983
QY 990 ELDELKAKYPMFTVNMISTGKPIDPSTKIKOWEAKYFNGTNVLGKAGYVLSDDATGKY 1049
Db 984 FLEELQAKYPELFAKQISTGVMPDPSVKIKOMSAKYFNGTNILGRGAGYVLDQATNTY 1043
QY 1050 FTWNENGQ--FLPASFTGDQNAKTGYFDGTGMAYYSTSGNKAVNSFIYEGGHYYFYDKD 1107
Db 1044 FNISDNKEINFLPKTLL-NQDSQVGFSDYDGKVVYVYSTSGYQAKNTFISEGDKWYFDNN 1102
QY 1108 GHMVTGSYKAEDGNDYIFLPGNIQMRDAIYODAQNSYVYVYGTGILYKGDNNWYVFPVDPNN 1167
Db 1103 GYMVTGA-QSINGVNYFLSNGLQRLDAILKNEDEGTAYAYGNDGRRYE-NGYVQFV--- 1156
QY 1168 ANKTVFRYEDANNVMAI GYRNMYGQTYFYDENGFOAKCOLLTDDKG-THYFEDENGAMAK 1226
Db 1157 --SGVNRHFN-NGEMSGLTVIDGQVQVDEMGYQAKKFVTTADGKIRIFYDKQSGNMVR 1213
QY 1227 NKXFV-NVGDDWYMDGNGNAVKGQVPMNQIILFYNPETGVQVKGQFITDAQGRTSYYDAN 1285
Db 1214 NREFIENECKWLYLGEDGAATVGSQTINGQHLYFR-ANGVQVKGFEVTDHGRISYYDGN 1272
QY 1286 SGALKSSGFFTPNGSDWYAEANGYVYKGFQVAENQDQWYFYDQTGQKAGKAQVGRD 1345
Db 1273 SGQDIRNR-----VRNAQGWYFYD-NNGYAVTGTARTINGQL 1309
QY 1346 LYFENDPSQVQVKGDFATDESNGNTSFVHGDNQDKVGGFFTTGNNWYAYADNNGNLVKGFQ 1405
Db 1310 LYFRA-NGVQVKGFEVTDYGRISYDNGSGDOIRNFRVRNAQGWYFYDNNNGYAV---- 1364
QY 1406 EIDGKWHYFDEVTGQQAAGALVNGQQLYFDVDSGTQVKGDFVTDGQGNYSYDVNSGDK 1465
Db 1365 -----TGARTINGQHLYFRA-NGVQVKGFEVTDHGRISYDNGSGDQ 1406

QY 1466 KVGFFFTGNAMYADGGNLAAGKRSIDNQDLYFDPATGKQVKGQVLSID-GRNYFDP 1524
 Db 1407 IRNFRVNAQGFYFNDNGYAVTGARTINGQHLF-RANGVQVKGFEVTDYGRISYD 1465

QY 1525 SGSGNMAK 1532
 Db 1466 ANSGVRV 1473

RESULT 13
 US-09-740-274-4
 ; Sequence 4, Application US/09740274
 ; Patent No. US20020031826A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nichols, Scott E.
 ; TITLE OF INVENTION: Glucan-containing Compositions and Paper
 ; FILE REFERENCE: 0357CRD
 ; CURRENT APPLICATION NUMBER: US/09740,274
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: 09/210,361
 ; PRIOR FILING DATE: 1998-12-11
 ; PRIOR APPLICATION NUMBER: 09/007,999
 ; PRIOR FILING DATE: 1998-01-16
 ; PRIOR APPLICATION NUMBER: 08/478,704
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: 09/009,620
 ; PRIOR FILING DATE: 1998-01-20
 ; PRIOR APPLICATION NUMBER: 08/485,243
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: 09/008,172
 ; PRIOR FILING DATE: 1998-01-16
 ; PRIOR APPLICATION NUMBER: 08/482,711
 ; PRIOR FILING DATE: 1995-06-07
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 1375
 ; TYPE: PRT
 ; ORGANISM: streptococcus mutans
 ; US-09-740-274-4

Query Match 38.4%; Score 3161.5; DB 3; Length 1375;
 Best Local Similarity 48.5%; Pred. No. 1.2e-158;

Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;

QY 1 MEKLLHYKLHKYKHWTTAVAS--IGLVSLVCA--GTVSAEDKVANDTTAQTAVGVDTG 56
 Db 1 MEKKVRFKLKVKRWVTVSIAVAVTLTSLGSLVKADSTDDRQQAFTESQASL----- 55

QY 57 QDOATTNDANTNT-TDPTDADQSANTNQDQAG-SDQSNQDQAKQDTANT-----DRNQ 108
 Db 56 ---VTTSEAKETLFTATDTSATSSQPTATVTDVNTNQSTNTTANTANFVVKPTTT 112

QY 109 ADNSQTDN-----NQATDQATS-----PATDGTSVQRDRANVAATAADQEG---QTAP 153
 Db 113 SEQAKTDNSDKITTSKAVNRLTATGKFPANNNTAHPKTVTDKIVPKPKGLKQPS 172

QY 154 SEQEKBAAL-SLDNVKLIDGKYVVOADGSYKQNFATVNGQMLYFSDTSGALSTSTYS 212
 Db 173 LSQDDTAALGNKVNKIRKVGKYYKKEDGTLOKQYALNTNGKTFPPD-ETGALSNNLTFS 231

QY 213 FSGQGTN--LVDDFFSHNKAYDSTAKSFELVNGYLTANSWYRPAGILRNGQWTEASNEND 270
 Db 232 KKGNTNNDNTNSFAQYNQVYSTDVANFEVDHYLFAESWYRPKYLLKDGKWTQSTKD 291

QY 271 LRPVLMSWPKDQTVAYVYNNKYLANSANETVNETSQVDLNMKEAQSIOTKIEQKITS 330
 Db 292 PRELLMTWPFQDQETQRYVYNNKAQLGIHQTYNT-ATSPQLNLAQAQTIQTKIEKITA 350

QY 331 NSTQWLRTAMEAFVAAQPKWNSTEN-FNKGHDLQCGALLYT-NSDLTPWANSYVRLNLR 388
 Db 351 KNTNMLRQTIISAVTKTQSAWNSDSEKPPD--DHLQKGALLYNNKSLTQSAWSNRYILNR 408

RESULT 14

US-10-383-930-35

; Sequence 35, Application US/10383930

; Publication No. US20040127400A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A

QY 389 TPTQDQGTK--KYFTGEGGEGYFLLSNVDNSNPVQVQABQLNOLHLYLNMWGDIVMGDKD 446
 Db 409 TPTNQTKDPRYTDRTTGGYFELLANDVDNSNPVQVQABQLNWLHFLMFGNIYANDPD 468

QY 447 ANFDGVRVDAVDNADLLQVSNYPKONTKYVTDSEANALAHISILSAWSLNDNQYNEDT 506
 Db 469 ANFDSIRVDAVDNVDADLLQIAGDYLKAAKGIHKNDKAAANDHLISLEAWSYNDPTYLHDD 528

QY 507 NGTALSIDNSSRLTSLAVLTQPGQRIDLSNLISSESUNKERANDTAYGDTIPTYSFVRAH 566
 Db 529 GDNMINMDNRLRUSLLYSLAKPLNQSGMPLITNSL-VNRTDDNAETAAPVSYSFTRAH 587

QY 567 DSEVQTVIAKIVKEKIDTNSDGYFTFLDQDKDAFKIYNEDMAKVNKYTHYNTPAAYALL 626
 Db 588 DSEVQDLIRNIIRTEINPNVVGYSFTTEEIKKAFIYNKOLLATEKYKTHYNTALSALL 647

QY 627 LSNMESVPRVYGDLYTDGQYMAKSPYDAITATMLQRIAYVSGQSSEVHKVNNQ 686
 Db 648 LTNKSSVPRVYGDLYTDGQYMAKSPYDAITATMLQRIAYVSGQSSEVHKVNNQ 706

QY 687 ILSSVRYGQDLSADDTQGTDLSTSLVTLVNSDNDPLDL-CGDSLTVNMGRAHANAYR 745
 Db 707 IITSVRYGKALKATDT-GDRITRTSGVAVIEGNNPSLRKASDRVNVNNGAAHKNQAYR 765

QY 746 PLILGTGQGVQSYLKDS-D-TNIVKYTDANGNLTFTADDIKGYSTVDMSGYLAVMVPVGA 804
 Db 766 PULLTTDNGIKAVHSDQEAAGLVRYTNDRGELIFTAADIKGYANPQVSGVYLVVPGAA 825

QY 805 DGQDVRVAADTNQKADGKSLKTSAAALDSQVYIEGFSNFPQFANNDADYTNKKIAENADFF 864
 Db 826 ADQDVRVAASTPSTDGKSVHQAALDSRVMPGFSNFPQAFATKKEEYTNVIAKNVDF 885

QY 865 KKLGITSPFMAQYVSATDGSFLDSIIQNGYASDRYDLAMSNNKYKSGKDDLANKAL 924
 Db 886 AEWGVTDFEMAPQYVSSDTGSLDSVIQNGYAFETDYLGISPKNYKGTADDLVKAIKAL 945

QY 925 HANGIOAIADWVPDQIYQLPGEVVTAKRTNSYGNPTFDAYINNALLYATNTKSSGSDYQA 984
 Db 946 HSKGIKVMADWVPDQMYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGKSGKQQA 1005

QY 985 QYGGAFDELKAKYPMFTVMNISTGKPIDPSTKIKOWEAKYPNGTNVLGKAGYVLSDD 1044
 Db 1006 KYGGAFLLELQAKYPLFARKQISTGVPMDPSVKIKQWSAKYPNGTNILGRGAGYVLDQ 1065

QY 1045 ATGKYFTVNEGDPLPASFT----GDONAKTFFYDGTGMAYYSTSGNKAIVSFIYEGGH 1100
 Db 1066 ATNTYFSLVSDNTFLPKSLVNPNNHGTSSSVTLGVPDGGKGYVYVYSTSGNQAFAISLGN 1125

QY 1101 YYYFDKDHMTGSKYKAEEDNDYFLPNGIQMRDAIYQDAQGSNYYGRTGILYK----- 1155
 Db 1126 WYFFDNGNGYVMTGA-QSINGANYFSLNGIQLNALLYDNGNKVLSYTGNDGRYENGYYL 1184

QY 1156 -GDNWYFVDPNNANKTVRYFDANNVMAIGYRNMYGQTYFYFDENGFOAKGOLLTDKG- 1213
 Db 1185 FQOQW-----RYFQ-NGIMAVGLTRVHGAQYVFDASGFOAKGQFITADGK 1229

QY 1214 THYFEDNCGAMAKNFV-NVGDDWYMDGNGHNAVKQYVPNNQILYFNBPETGVQVKGQFI 1272
 Db 1230 LRYFDRDSNQISNRPVRNSKGEWFLDHNGAVTGTVTFTNGQRLYFKP-NGVOAKGEFI 1288

QY 1273 TDAQRTSYDDANSALKSSGFFTPNGSDWY-YAENGYYVYKPKQV 1317
 Db 1289 RDANGYLRYDPSNGNEVRNFRVNSKGEWFLDHNGIAVTGARVV 1334

; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PR1
; ORGANISM: Streptococcus mutans
; US-10-383-930-35

Query Match 38.4%; Score 3161.5; DB 4; Length 1375;
Best Local Similarity 48.5%; Pred. No. 1.2e-158;
Matches 662; Conservative 223; Mismatches 400; Indels 81; Gaps 31;

QY 1 MEKKLHYKLVKVKHWTIIVAS--IGLSVLGA--GTVSAEDKVANDTTAQAATVGDVG 56
DB 1 MEKKVRFKLRKVKRWVTVSIASAVTLTSLGSLVKADSTDDRQOAVTESQASL----- 55
QY 57 QDOATTNDANTNT-TDITDADQSANTNQOAG--SDQSNQOQAKQDTANT-----DRNQ 108
DB 56 ---VTTSEAAKETLTATDTSTATSATSQPTATVTDVNSTTQSTNTTANTANFVVKPTTT 112
QY 109 ADNSQTDN-----NOATDOATS-----PATDGTSGVORRDAANVATAADQEG---QTAP 153
DB 113 SEQAKTDNSDKIITTSKANRLTATGKFPANNHTAHPKVTVDKIPIKPKLQKQPS 172
QY 154 SEQEKAAL-SLDNVLIDGKYVVOAGDSYKKNFAITVNGQMLYDSDTSGALSSSTYS 212
DB 173 LSQDDIAALGNVKNIRKNGKYYYYKEDGTQKNYALNINGKTFPFD-ETGALSNNTPS 231
QY 213 FSGQTN--LVDDFSSHKAYDSTAKSPFLVNGYLTANSWYRPAIIRNGQITWEASNE 270
DB 232 KKGNIITNDNTNSFAQYNQVYSTDVANFEHVDHYLTAESWYRPKYILKDKGTWTQSTKD 291
QY 271 LREVLMSWPKDQTVAYVYNNMKYLSANETVNETSVDNLNKEAQSIQTKIEOKITSD 330
DB 292 PRELLMTWFPDQETQRYVYNNMAQLGIHQTYNT-ATSPLOQLMLAAQTQTKIEEKITAE 350
QY 331 NSTQWLRTAMEAFVAAQPKWNMSTEN-FNKGDLQGGALLYT-NSDLTPMANSDYRLNLR 388
DB 351 KNTNMLRQITISAPVKTSQAWNSDEKFPD--DHLQKGLLYSNNSKLTQANSNYRILNR 408
QY 389 TPTQOQGTGK--KYFTGGGGYEFLLSDVDNSNPVVQAEOLNQLHYLNNWGDIVMGDKD 446
DB 409 TPTNQTGKDPRTYDRTTGGYEFLLANDVDNSNPVVQAEQLNMLHFLMFGNIYANDPD 468
QY 447 ANFDGVRVDADVNNADLLQVYNSYKDNVYKVTDSANALAHISILEAWSLNDNQYNEPT 506
DB 469 ANFDSIRVADVNDVADDLQIADGYLKAAGIHKNDKAANDHLISILEAWSYNDTPYLHDD 528
QY 507 NGTALSIDNSSRLTSLAVLTQPGQRIDLSNLISSEVKNKERANDTYAGDTIPTYSVRAH 566
DB 529 GDMNINNRLRLSLLYSLAKPLNQRSGMPLITNSL-VNRTDDNAETAAPVSYSFIRAH 587
QY 567 DSEVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAFKYINEMAKVNKYTHYNTIPAAVALL 626
DB 588 DSEVQDLIRIIRTEINPNVNVGSPTTEIKKAFEYINKDILLATEKKYKTHYNTALSAYLL 647
QY 627 LSNMESVPRVYVYGDLYTDDQYMAKSPYYDAITMLQGRIAVYSGQSEVHKVGNNGQ 686
DB 648 LTNKSSVPRVYVYGDYMTDDQYMAKHTINYEALETLLKARIKTVSGQAMRNOQV-GNSE 706
QY 687 ILSSVRYGQDLMSADDTQGTDLRSRTSLGVLTVSNDPNLDL--GGDSLTVNNGRAHQAQYR 745
DB 707 IITSVRYGKALKATDT-GDRTTTSVGAIVIEGNPNPSRLKASDRVVVNNMGAHKKQAYR 765
QY 746 PLILGTGKGQVQYLKQSD--TNIYKYTDANGNLFTTADDINKGYSTVDSMGVLAWVVPVGA 804

DB 766 PLLITTDNGIKAYHSDQEAAGLVRYTNDREGILFTAADIKGYANPQVSGYLVVWVPGAA 825
QY 805 DQGDVVRVAADTNQKADGSKLTKTSAALDSQVIYEGVSFNFODFANNADADYTNKKIAENADPF 864
DB 826 ADQDVRVAASTAPSTDGKSVHQNAALDSRVMEGFSNFOAFATKKEEYTNVVIKAVDKF 885
QY 865 KKLGLITSFEMAPQYVSATDGSFLDSIQNGYAFSDRYDLAMSKNNKYGSKDDLANKALKAL 924
DB 886 AEWGVTDFEMAPQYVSSTDGSFLDSVIQNGYAFTRDYLGISKPNKYGTADDLVKAIKAL 945
QY 925 HANGIQATADWVPDQIYOLPGEEVVTAKRTNSYGNPTFDAYINNALLYATNTKSSGSDYQA 984
DB 946 HSKGILKWADWVPDQMYALPEKEVVTATRVKYGTFVAGSQIKNTLYVVDGKSSGKDDQA 1005
QY 985 QYGGAFLELQAKYPMFTVMNMISTGKPIDPSTKIKQWEAKYFNGTNVLGKAGYVLSDD 1044
DB 1006 KYGGAFLLELOAKYPELPARKQISTGVPMDPSVKIKQWSAKYFNGTNILGRGAGYVLKQD 1065
QY 1045 ATGKYFTVNEGDPLPASFT-----GDQNAKTGFYDGTGTMAYYSTSGKNKAVNSFIYEGGH 1100
DB 1066 ATNTYFSLVSDNTFLPKSLVNPNNHGTSSSVTGLVFDGKGYYVYSTSGNQAKNAFTSLGN 1125
QY 1101 YYYFDKGHMTGSKYKAEDGNDYFPLNGIQMRDAIYODAQGNSYYVYGRGTGLYK----- 1155
DB 1126 WYTFDNGWNVYGA-QSINGANYFSLNGIQIRNALYDNGNKVLSYNGDGRYENGYYL 1184
QY 1156 -GDNWYFPVDPNNANKTVRYFDANNVMAIGYRNMYGQTYFYFDENGFOAKGQLLTDKG- 1213
DB 1185 FGOQW-----RYFQ-NGIMAVGLTRVHGAVQYFDASGFOAKGQFITTADGK 1229
QY 1214 THYFEDNGAMAKNFV-NVGDDWYMDGNGNAVKQYVNNQILLYFNSETGVQVKGQFI 1272
DB 1230 LRYFDRSDNQISNRFVRNSKGWFLDHNGVAVTGTVTFNGQRLYFKP-NGVQAKGEFI 1288
QY 1273 TDAQRTSYDANSALKSSGFTTPNGSDWY-VAENGYVYVKGFKQV 1317
DB 1289 DRANGVLYRYPNSGNEVRNFRVNRKSGWFLDHNGIATGARVV 1334

RESULT 15

; US-10-797-821-35
; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PR1
; ORGANISM: Streptococcus mutans
; US-10-797-821-35

Query Match 38.4%; Score 3161.5; DB 5; Length 1375;
Best Local Similarity 48.5%; Pred. No. 1.2e-158;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 8.84317 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-38
Perfect score: 8237
Sequence: 1 MEKKLHYKLHKVKKHWTTA.....FVRIGQWYFGNDGAATNL 1554

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 1312538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New.*
1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pdb.*
2: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pdb.*
3: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pdb.*
4: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pdb.*
5: /cgn2_6/prodata/1/pubpaa/US05_NEW_PUB.pdb.*
6: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pdb.*
7: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pdb.*
8: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pdb.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	545.5	6.6	2710	US-11-051-453-41	Sequence 41, Appl
2	443.5	5.4	2367	US-11-051-453-42	Sequence 42, Appl
3	363	4.4	1992	US-11-013-759-3	Sequence 3, Appl
4	363	4.4	1992	US-11-013-759-13	Sequence 13, Appl
5	363	4.4	2047	US-11-013-759-4	Sequence 4, Appl
6	363	4.4	2047	US-11-013-759-7	Sequence 7, Appl
7	341.5	4.1	2053	US-11-013-759-9	Sequence 9, Appl
8	306.5	3.7	2314	US-11-013-759-11	Sequence 11, Appl
9	302.5	3.7	396	US-11-022-562-228	Sequence 228, Appl
10	290	3.5	726	US-11-052-554A-217	Sequence 217, Appl
11	286.5	3.5	1647	US-11-052-554A-260	Sequence 260, Appl
12	276.5	3.4	2399	US-11-052-554A-92	Sequence 92, Appl
13	267	3.2	338	US-11-052-554A-228	Sequence 228, Appl
14	267	3.2	1155	US-10-793-626-1780	Sequence 1780, Ad
15	265	3.2	1767	US-11-052-554A-372	Sequence 372, Appl
16	263	3.2	3194	US-11-052-554A-90	Sequence 90, Appl
17	261.5	3.2	744	US-10-873-528-184	Sequence 184, Appl
18	258	3.1	2902	US-11-052-554A-91	Sequence 91, Appl
19	256.5	3.1	677	US-10-873-528-155	Sequence 155, Appl
20	256	3.1	619	US-11-052-554A-229	Sequence 229, Appl
21	256	3.1	701	US-11-052-554A-231	Sequence 231, Appl
22	255	3.1	658	US-10-873-528-17	Sequence 17, Appl
23	251	3.0	5291	US-11-052-554A-281	Sequence 281, Appl
24	249	3.0	1588	US-11-052-554A-280	Sequence 280, Appl
25	246.5	3.0	690	US-11-052-554A-232	Sequence 232, Appl

26	245.5	3.0	1290	6	US-10-485-517-141	Sequence 141, App
27	241.5	2.9	485	6	US-10-630-203-2	Sequence 2, Appli
28	241.5	2.9	485	7	US-11-103-037-1	Sequence 1, Appli
29	241.5	2.9	485	7	US-11-195-538-1	Sequence 1, Appli
30	240	2.9	332	6	US-10-873-528-190	Sequence 190, App
31	235	2.9	1571	7	US-11-052-554A-2	Sequence 2, Appli
32	228.5	2.8	2551	7	US-11-052-554A-368	Sequence 368, App
33	227.5	2.8	5024	6	US-10-793-626-2964	Sequence 2964, Ap
34	226.5	2.7	1268	7	US-11-052-554A-1	Sequence 1, Appli
35	224.5	2.7	1448	6	US-10-485-517-212	Sequence 212, App
36	223	2.7	340	6	US-10-873-528-188	Sequence 188, App
37	220	2.7	693	6	US-10-873-528-185	Sequence 185, App
38	219.5	2.7	1345	7	US-11-052-554A-282	Sequence 282, App
39	216	2.6	1327	7	US-11-052-554A-13	Sequence 13, Appl
40	216	2.6	1889	7	US-11-102-476-46	Sequence 46, Appl
41	212	2.6	1263	6	US-10-485-517-127	Sequence 127, App
42	210	2.5	1461	7	US-11-052-554A-283	Sequence 283, App
43	209	2.5	627	6	US-10-873-528-191	Sequence 191, App
44	208	2.5	1484	6	US-10-517-939-74	Sequence 74, Appl
45	206.5	2.5	567	6	US-10-485-517-216	Sequence 216, App

ALIGNMENTS

RESULT 1

US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-FEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 6.6%; Score 545.5; DB 7; Length 2710;
Best Local Similarity 20.1%; Pred. No. 1.2e-20;
Matches 380; Conservative 226; Mismatches 624; Indels 661; Gaps 91;

QY	35	VSADKAVNDTTAQATVGTGQDQATTNDANTTTDTTADOSANTNQDQAG-----S 88
Db	913	VETEKEIFSKYSEHITKEITKNIIIT-DVGNLLDNLQDLHTSOVTLNAAFFTQSLL 971
QY	89	DQSNQD-----QAQDQDANT--DRNQADNSQTDNNNAQDQATSPATDGT 131
Db	972	DYSSNKQVLDLSTSVKVLQAFLSTGLTYDLSQLVNLSNAVNDINVLPTTEGI 1031
QY	132	SVQRR--DAANVATA----ADQEGTAPSEQKSAALSNDVNL-IDGKYVYVQADGSYK 184
Db	1032	PIVSTILDGINLGAIAKELLDEHPLKLEAKVGLVAINMNSLSAATVASTIVGIGAEV 1091

RESULT 4
 US-11-013-759-13
 ; Sequence 13, Application US/11013759
 ; Publication No. US20050249747A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Loomore, Sheena M.
 ; APPLICANT: Sasaki, Ken
 ; APPLICANT: Yang, Yan Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
 ; TITLE OF INVENTION: PROTEIN OF MORAXELLA
 ; FILE REFERENCE: 1038-921MIS:jb
 ; CURRENT APPLICATION NUMBER: US/11/013,759
 ; CURRENT FILING DATE: 2004-12-16
 ; PRIOR APPLICATION NUMBER: US/09/361,619
 ; PRIOR FILING DATE: 1999-07-27
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 13
 ; LENGTH: 1992
 ; TYPE: PRT
 ; ORGANISM: Moraxella catarrhalis
 US-11-013-759-13
 Query Match 4.4%; Score 363; DB 7; Length 1992;
 Best Local Similarity 20.6%; Pred.No. 2.5e-11;
 Matches 377; Conservative 196; Mismatches 682; Indels 578; Gaps 86;
 QY 18 TIAVASIGLVSLVGAGTVAEDKAVNDTTAQTAVGDTVGQDQAVTNDANTN-----TTDTD 73
 DB 196 TIAIGSDATSSSGLAALGA-----GTRAQLQSGIALGQGVVTSQDNNRPAFTPTQ 249
 QY 74 TADQSAN--TNQQAQSGDSGNQDQAKQDQTANTDRQADNSQTDNNQATDQATSPATDGTG 132
 DB 250 ALDPKFAQNTNKGKAG--PLSIGSNSIKRKIIINV-----GAG 283
 QY 133 VORRDAANVAT-----AADQEGQTAPSEQSAALSLDNVKLDGKY----- 174
 DB 284 VNKTDAVNVQAEAVVQKAKERRITPQGDNDSTVKIGLDNTLTIKGAETNALTDNNIG 343
 QY 175 YVYQADGS--YKKNFAITVNGQMLYFSDPTGALSSTSY-----SFSQGTNLVDD-----FSSH 227
 DB 344 VVKEADNSGLKVKLAKTLNN--LLEVNTTLNATTTVKVSGSSSTTAEELSDSLFTQP 400
 QY 228 NKAYDSTAKSFELVNGYLTAWSRPAQILNGQDWEASNENLRLPVLMSWWPDKQTOVA 287
 DB 401 NTGSOQSTKTVYGVNGVKFTN-----NAETTAAGTTRITRDKIGFARDGVDYDEK 450
 QY 288 YVYNNV--KVLNANETEVNNETSQVDLNKEAQSIQTKIEQKITSQNSTQWLRTAMEAFVA 346
 DB 451 QAPYLDKQLKVGSAVITIDGIDAGNKKISNL-----AKGSSANDA-----VTIEQLKAA 501
 QY 347 QPKNNMSTENFNKGDHLOGGALLYTNLSLTPWANS-----DYRL--LNRTPQDQGTQKY 399
 DB 502 KPTLN-----AGAGISVTPTEISVDKSGNVTAFTYNIQVKTTELNSDGTSDK 549
 QY 400 FTEGEGGVEFLLSNDVNSHPVQAEOL--NQLHYLNNWGDIV-----GDKANFD 450
 DB 550 FSVKSGS-----TNNSLVTAETHASLYLNEVARTADSALQSFYVKEEDDDANA 598
 QY 451 GYRVDAVDNVNADLLQVYSNYFKONYKYVDEANAL-----AHISILEAWSLNDNQYN 503
 DB 599 TVAKOTTKNAGA--VSILKLGKGLVATVKDGTVPFGLSQDSGLTIGKSTLNNGLTV 656
 QY 504 EDTN-----GTALSIDNSRLT----- 520
 DB 657 KDTNEIQIQVANGIKFTNVNGSNPCTGIA--NTARITRDKIGFAGSDGAVTNKPYLDQD 714
 QY 521 -----1-----SLAVLTQPGQRIIDLNSLSESVNKEANDTAYG 554
 DB 715 KLQGVNKITWGINAGKATITGLSPTLPSIADQSSRNIELGTIQD---KDKSNAASIN 771

Db 1660 AGQTNVLTNNPAAIRDRNEQGRFFHVNDGNEP-----VQGRNGIDSSASGKHSVAIG 1715
Qy 1530 -----MAKNRFVRIGDQWYFNDGAAAT 1552
Db 1716 FOAKADGEAAVAIGRQTQAGNQSIAGIDNAQAT 1748
RESULT 5
US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasami, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-4
Query Match 4.4%; Score 363; DB 7; Length 2047;
Best Local Similarity 20.6%; Pred. No. 2.6e-11;
Matches 377; Conservative 196; Mismatches 682; Indels 578; Gaps 86;
Qy 18 TIAVASIGLSVLGAGTVSAEDKVANDTTAAQATVGVDTGQDQATTNDANTN-----TTDTD 73
Db 251 TIAIGSDATSSSLGATGALG-----GTRAOLGSGALGQGSVVVTQSDNNRSPAYTPNTQ 304
Qy 74 TADQSAN-TNQDQAGSDQSNQDQAKQDQTANTDRNQADNSQTDNNQATQATSPATDGTG 132
Db 305 ALDPKQATNNTKAG-PLSIGSNSIKRKIINV-----GAG 338
Qy 133 VQRDDAANVAT-----AADQEGQAPSEQKSAALSDNVKLDGKY----- 174
Db 339 VNKTDVAVNAQLEAVVKAERITFGQDDNSDVKIGLDNTLTIRGGAETNALTDNNIG 398
Qy 175 YYVOADGS-VKKNFALTNGQMLYFSDTGALGSTSTY---SFSQGTNLVDD---FSSH 227
Db 399 VVKEADNSGLKVKLAKTLNN---LTEVNTTTLNATTTVKVGSSTTAELLSLSLTFTQP 455
Qy 228 NKAYDSTAKSFELVNGYLTANSWYRPAIGLRNQGTWEASNENDLRPVLMWSWPDKDTQVA 287
Db 456 NTCQSQSTKTVGVGVKFTN-----NAETTAAGTTRITRDKIGFARGDGVDEK 505
Qy 288 YVNMN-KYLSANETEVNTSETQVDLNKEAQSIQTQKIEKISDNTQMLRTAMEAFVA 346
Db 506 QAPYLDKKQLKVGSAITIDNGIDAGNKKISNL-----AKGSSANDA---VTIEQLKAA 556
Qy 347 QPKWNSTENFNKGDHLOGGALLYTNSDLTPWANS-----DYRL-LNRPTQODGCKKY 399
Db 557 KPTLN-----AGAGISVTPTEISVDKSGNVVTAPTNIGVKTITELNSDGTSDK 604
Qy 400 FTEGGEGGYEFLLSNDVDSNPVVQAEQL-NQLHYLMNWGDIVW-----GDKDANFD 450
Db 605 FSVKSG-----TNSLVTAHLASYLNEVNRADTSALQSFVTKKEEDDDANAI 653
Qy 451 GVRVADVNNADLLQVYSYFNKDYKVTQSEANAL-----AHISILEAWSLNQN 503
Db 654 TVAKDITTKNAGA--VSILKLGKNGLTVAATKDDGTVTFGLSQDSGLTIGKSTLNDGLTV 711
Qy 504 EDTN-----GTALSIDNSSRLT----- 520

Db 712 KUTNEQIQVGANGIKFTNVGNSPNGTGIA--NTAITRDKIGFAGSDGAVTDNKPYLDDQ 769
Qy 521 -----SLAVLTQPGQRIIDLNLNLSVESVKNERANDTAYG 554
Db 770 KLVGVNKITNTGINAGGKAITGLSPTLPSIAIQSSRNIELGNTIQD--KDKSNAASIN 826
Qy 555 DTIPYISFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAFKIYNEDMAKNKY 614
Db 827 DILNTGFNLKNNNPIDFVSTYDI VDFANGNATTATVTVTHDTANKTSKVYD--VNVDDTT 884
Qy 615 THYNIPAAVALLSNMESVPRVYVYGLYTDGQYMAKSPYYDAIATMLQGRILAYVSGQ 674
Db 885 IHLT-----GTDDNKLGVKT-----TKLNKTA----- 908
Qy 675 SEEVHKVNGNNOILSVRYGQD--LMSADD-----TQGT-DLRSRTSLGLVTL 717
Db 909 -----NGTATNFNVNSSDEDAVLNAKDIAENLTLAKEIHHTKGTADTALQTFVVK 961
Qy 718 VSNDPNLDLGGDSLTVNMGRAHANQAYRLLIIGTKGVOSYKSDSTNIVKYTDANGNLT 777
Db 962 VDENNNAD--DANAITVQKNNQVNTLTLLKGENL-----NI--KTDRNGVT 1007
Qy 778 F---TADDIK-GYSTVDMSGYLAVVPVGAKD---QGD--VRVAADTNQKADGKSLKTSAA 829
Db 1008 FGINTTSLGKAKSKTLN-DGGLSINKPTGSEQIQVGADGVKFAKVNNGVVGAGIDGTT 1066
Qy 830 LDSQVIYEGFSNFPQDPAANDADYTNKKIAB--NADFFKXGLITSEMAPQVYSATDGSF 886
Db 1067 ITRDEI--GFTG---TNGSLDKSPHLSKDGINKAGGKKTNIQSGEIAQNSHDAVTGGK 1120
Qy 887 LDSIQNGYAFSDRYDLAMSKNKGSKDDLANALKALH-----ANGIOAI 932
Db 1121 I-----YDLKTELENKISS--TAKTAQNSLHEFSVADEQGNFTVSNPYSS 1165
Qy 933 ADWVPDQIYQLPGEVVTAK-----RTNSYGNPTFDAYINNA----- 969
Db 1166 DTSKTSDVITPAGENGITTKVNGKVVVRIGIDTKGLTTPKLTGVNNGKGIIVDSQNGN 1225
Qy 970 -----LYATNTKSSGSDYQAYGGAFLDELAKAYPDMFTVMNIMSTGKPIDPSTKIKOW 1022
Db 1226 TITGLSNTLANVTNDKGSVRTTEQGNIIKDEKTRAASI--VDVLSA----- 1270
Qy 1023 EAKYFNGTIVLKGAGYVLSDDATGKYFTVN--ENGDFLPASFTGDNQATG--FYD--- 1076
Db 1271 -----GFNLOGGEAV---DFVSTYDTVNFADGNATTAKVTYDDTSTKSVVYDVNVD 1320
Qy 1077 -----GTGMAYYSTSGNKAVNSFIYEG---GHYYFPDKDGHMV 1111
Db 1321 DTTIEVKDKKLGKVTTLTSTGTGANKFALS--NQATGDALVKASDIVAHLNTLSGDIQTA 1379
Qy 1112 TGSYKAEDGNDYFPLPNGIQMRDAIYQDAQNSYYVYGRGILY---KGNWYFPVDPNNA 1168
Db 1380 KGASQANN-----SAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKEVAK 1425
Qy 1169 NKTVPYFPDANNVMA-IGYRNMYGQTYYPDENGFOAKGQLLTTD----KGTHYFDEDNGA 1223
Db 1426 DKLVAQAQPTPDGTLAQMNKVSINKEQVNDAN---KKGINEDNFAVFKLEKASADN-- 1479
Qy 1224 MAKNFVNVYDQWYMDGNGNAVKQYPVNNQILVFNPTGVQVK--GQFITDAQRTSY 1281
Db 1480 KTNAAVTVGD-----LNAV-AQTP-----LTFAGDTGTTAKLGETLTIKGQOT-- 1523
Qy 1282 YDANGALKSSGFFTPNGSDWYEAENGYYVYKGFQVAENQDQWYYPDQTTGKOAKGAHV 1341
Db 1524 -DTNK-----LTDNNIGVVAGTDGFTVKLAKDLT-----NLNSVNAAGTKI 1563
Qy 1342 DGRDLYFPNDSGVQVKGDFATDESNGTSFYHG---DNGDKVYVGFPTTGNNAWYADNNG 1398
Db 1564 DDKGSFVDSG--QAK-----ANTPVLANSGLDGLGKVI-----S 1597
Qy 1399 NLVKGFQETD-GKWYHFDEVTGQAKGAA-----LVNGQOLYFVDSGIO-----VK 1444

Db 1598 NVGKGTXTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSRRVTIK 1657

Qy 1445 GDFVTDQGGNTSYDVNSG-----DKVANGFFTTGD--NAW-----YYA 1481

Db 1658 AGTVLGGKGNNDTEKLATGGIQGVDRDGN---ANGDLSNVWVKTKQDGSKALLATYNA 1714

Qy 1482 DGO-----GNLAKGRKSIDNQDL-YFDPATGKQVKQLVSDIGRNYFYDFSGSN----- 1529

Db 1715 AGQNTVLTNNPAEADIRNEQGRFFHFVNDGNQEP-----VVQGRNGIDSSASGKSHVAIG 1770

Qy 1530 -----MAKNRFVRIGDQWIFGNDGAAT 1552

Db 1771 FOAKADGEAAVAGROTAQAGNSIAIGDNAQAT 1803

RESULT 6

US-11-013-759-7

; Sequence 7, Application US/11013759

; Publication No. US20050249747A1

; GENERAL INFORMATION:

; APPLICANT: Loomore, Sheena M.

; APPLICANT: Sasaki, Ken

; APPLICANT: Yang, Yan Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE

; TITLE OF INVENTION: PROTEIN OF MORAXELLA

; FILE REFERENCE: 1038-921MIS:jb

; CURRENT APPLICATION NUMBER: US/11/013,759

; CURRENT FILING DATE: 2004-12-16

; PRIOR APPLICATION NUMBER: US/09/361,619

; PRIOR FILING DATE: 1999-07-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 7

; LENGTH: 2047

; TYPE: PRT

; ORGANISM: Moraxella catarrhalis

US-11-013-759-7

Query Match 4.4%; Score 363; DB 7; Length 2047;

Best Local Similarity 20.6%; Pred. No. 2.6e-11;

Matches 377; Conservative 196; Mismatches 682; Indels 578; Gaps 86;

Qy 18 TIASIGLSVLVAGCTVSAEDKAVNDTTAAQTAVGVDTGQDQATTNDANTN-----TTDTD 73

Db 251 TIAIGSDATSSSLGALGA-----GTRAQGSLAGGVVVTQSDNNSRPAYTPNQ 304

Qy 74 TADQSAN-TNQDAGSDQSNQDQAKQDPTANTDRNOADNSQTDNNQATQATS PATDGT 132

Db 305 ALDPKFOATNNTKAG-PLSIGNSIKRKIINV-----GAG 338

Qy 133 VORRDAANVAT-----AADQSGQAPSEQKSAALSIDNVKLIIDGKY----- 174

Db 339 VNKTDAVVAQLEAVVWAKERRITFGQDNDSTVDVKGIDNTLTIKGAETNALTNNIG 398

Qy 175 YVVOADGS-YKKNFAITVNGQMLYFSDTCALSSSTSY-----SFSQGTTLNVD-----FSSH 227

Db 399 VVKEADNSGLVKLAKTLN---LTVNTTLNATTVKVGSSSTTAEILLSLFTPTQ 455

Qy 228 NKAYDSTAKSFELVNGYLTFANSWYRPAGILRNGQTWEASNEENDLRPVLMSWMPDKDTQVA 287

Db 456 NRGQSSTKTVGVNGVVKFTN-----NAETTAAGITTRITRKIGFARGDGVDEK 505

Qy 288 VYVYMN-KYLSANEIEVTNETSQVDLANKAQSLOTKIEOKITSDNSTQWLRTAMEAFVAA 346

Db 506 QAPYLDKKQLKVGVSVAITIDNGIDAGNKKISNL-----AKGSSANDA-----VTIEQLKAA 556

Qy 347 QPKWNNSTENFNKGDLQGGALLYTNSDLTPWANS-----DYRL-LNRTPQDQTKKY 399

Db 557 KETLN-----AGAGISVTPTEISVDAKSGNVTPATYNTGVKVTIELNSDGTSDK 604

Qy 400 FTEGEGGGYEFLLSNDVDNSNPFVQAEQL-NOLHYLMNMGDIVM-----GDKDANFD 450

Db 605 FSVKSGS-----TNNSLVTAHFLASLYLNEVNRRTADSALQSFVTKBEDDDANAI 653

Qy 451 GVRVDAVDNVADLLQVYSNYFKONYKVTDSEANAL-----AHISILBAMSLNDNQYN 503

Db 654 TVAKDTTKAGA--VSILKLGKNGGLTVATKDGTVTFGLSQDSGLTIGKSTLNDGLTV 711

Qy 504 EDYN-----GTALSIDNSSRLT----- 520

Db 712 KDTNEQIQVGANGIKFTNVNGSNPGTICIA--NTARITRDKIGFAGSDGAVDTNKPFLYDQD 769

Qy 521 -----SLAVLTQPGQORIDLNLISSEVNVKERANDTAYG 554

Db 770 KLQGVNWKITNTGINAGKAITGLSPISIADQSSRNIELGNTIQD-----KDKSNAASIN 826

Qy 555 DTIPTYSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQKDAFKIVNEDMAKVNKTY 614

Db 827 DILANTGFNLKNNNPIDFVSTYDIVDFANGNATATVTHDTANKTSKVYD--VVUDDTT 884

Qy 615 THYNIPAAVALLLSNMESVPRVYVYDLYDDGQYMAKSPYYDAIATMLQGRYAYVSGGQ 674

Db 885 IHLT-----GTDDNKKLGVKT-----TKLNKTS----- 908

Qy 675 SEEVHKVNGNNQILSSVRYQD--LMSADD-----TOGT-DLSTSGLVTL 717

Db 909 -----NGTATNFNVNSSDEDALYNKADIAENLNTLAKEIHTTRGTADTALQTFVVK 961

Qy 718 VSNDPNLDLGGDSLTVNMGRANQAQYRPLILGTQGVQSYLQSDTNI VKYTDANGNLT 777

Db 962 VDENNAD--DANAITVGKXANNQVNTLTXENGL-----NI--KTDKNGT 1007

Qy 778 P---TADDIK-GYSTVDMSGYLAVVYVGAQD---QD-VRVAADTNQKADGKSLKTSAA 829

Db 1008 FGINTTSGLKAGKSTLN-DGSLSIKNPTGSEQIQVGADGVKPAKVNNGVVGAGIDGTT 1066

Qy 830 LDSQVIYEGFSNFPQDPANNDADYTNKCIAB---NADFFKLGLTSEMAPQYVSATDGSF 886

Db 1067 ITRDEI--GFTG---TNGSLDKSPHLSKOGINAGGKTIITNIOGSEIAQNSHDAVTGK 1120

Qy 887 LDSIIQNGVAFSDRYDLAMS KKNYGS KDDLANALKAH-----ANGLOAI 932

Db 1121 I-----YDLKTELENKISS--TAKTAQNSLHEFSVADQGNFTVSNPYSSY 1165

Qy 933 ADMVPDQIYQLPEEYVTA-----RTNSYGNPTFDAYINNA----- 969

Db 1166 DISKTSVITFAGENGITTKVNGVVRVIGDQTKGLTTPKLTGVNNGKGIVIDSQNGQ 1225

Qy 970 -----LYATVTKSSGSDYQAGAFDELKAKYDPMPTVMIMSTGKPIDPSTIKQW 1022

Db 1226 TITGLSNTLANVTNDKGSVRTTEQGNIIKDEDKTRAASI--VDVLSA----- 1270

Qy 1023 EAKYFNGTNVLGKAGYVLSDDATGKYFTVN-ENGDFLPASFTGQDQNAKTG-FYYP----- 1076

Db 1271 -----GFLQNGEAV-----DFVSTYDTVNFADGNATTAKVYDTSKTSKVYVDNV 1320

Qy 1077 -----GTGMAYYSTGSKAVNSFIYEG-----GHVYFPDKGHMV 1111

Db 1321 DTTIEVKDKKGLVKVTKTLTSTGTGANKFALS-NQATGDALVKASDIVAHLNTLSGDIQTA 1379

Qy 1112 TGSYKAEDGNDYFPLPNGIQMRDAIYQDAQNSYYVYGRGTIILY---KGDNWPFPVDPNNA 1168

Db 1380 KGASQANN-----SAGYVDADGNKVIYDSTDNKYVQAKNDGTVDKTEVAK 1425

Qy 1169 NKTVPYFPDANNVMA-IGYRNMYGQTYFDENGFQAKQLLTD-----KGTHYFDENGA 1223

Db 1426 DKLVAQAQPTPGDTLQAMNVKSVINKEQVNDAN---KKGINEDNAFVKLEKASDN-- 1479

Qy 1224 MAKNFVNVGDDWYMDGNNAVKQYPVNNQILYFNPNBTGVQVK--GOFITDAQORTSY 1281

Db 1480 KTKNAAVTGD-----LNAV-AQTP-----LTFAGDTGTAKKGLBTLIKGGQT-- 1523

Qy 1282 YDANSALKSGSFFTPNGSDMYAENGYYVYKGFQVAENQDQWYDPDTTGGQAKGAKV 1341

Db 1524 -DTNK-----LTDNNIGVVAGTDGFTVKLAKDLT-----NLNSVNAAGTKI 1563

Qy	1342	DGRDLYPNDSGVQVKQDFATDESCNTSFYHG---	DNGDKVVGFFTTGNNAWYYADNNG	1398
Db	1564	DDKGVSVFVDSG-QAK-	---ANTFVLSANGLDLGGKYI-	1597
Qy	1399	NLVKGFQEID-KRWYHFDEVTGQAKAA-	---LVNGQQLYFDVDSGIC-	1444
Db	1598	NVGKGTDTDAANVOQLNEVRNLLGLGNAGNDNADNQVNIADI	KKDPNSGSSNRVVIK	1657
Qy	1445	GDFVTDGQGNFSYYDVNSG-	---DKKXNGFFTTGD--NAN	1481
Db	1658	AGTVLGGKGNNDTEKLTAGTIGVQVDKGN-	---ANGDLSNVWVKTKDGSKKALLATYNA	1714
Qy	1482	DGQ----	GNLAKGRKSIDNQLL-YFDPATGKQVKGLVSDGRNYFYFDSGSGN-	1529
Db	1715	AGQTNLYLTNPAPAEIADRINEGGIRFFHVDNGQEP----	VVQGRNGIDSASSGKHSVAIG	1770
Qy	1530	-----MAKRVFVRIRIGQWIFYGNDGAAT	1552	
Db	1771	FOAKADGEAAVAIGROTQAGNOSTAIGDNGAQT	1803	

RESULT 7

```

US-11-013-759-9
; Sequence 9, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2053
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
; US-11-013-759-9

```

	Query Match	4.1%; Score 341.5; DB 7; Length 2053;
	Best Local Similarity	20.9%; Pred. No. 3.4e-10;
	Matches 380; Conservative 211; Mismatches 667; Indels 557; Gaps 95;	
Qy	21 VASIGLVSLVCAGTVAEDKIVANDTTAAQTGVGD-----TGODQATT-----NDANTN	68
Dd	269 VVNLGRGVALGFSGQLIDRRDNTDASAIVPLGKTLADQYKATRGDSSTDFPSIGNSNNN	328
Qy	69 TT-----DTDT-----ADOSANTNQAGSDQNNOQAQDPANTDRN	107
Dd	329 SSIRPKINVCAGRSDDTVANNVAQLKLVEELANRKITFKG-DGDNNNSVERGLGNTLTI	387
Qy	108 QADNSQTDNNQATPQATSPTDGTSVQRDDAANYATAADBOGOTAPSEQEKAALS--D	165
Dd	388 KGD-AQT-NALTEANIGVTWTDGNLKG-----VKLAKELTGLTSVSATNKITVSENTNN	438
Qy	166 NVKLIDGGYYVQADGSYKKNFALTIVNGOMLYPPSDTCALSSSTYSFSQ-----GTIN	219
Dd	439 NAELOSGLTFSPITGTIKTKDVSVSIDG--LKFTNDSINSIATKTTRITKKKIGFAGTND	496
Qy	220 LVDDFFSH-----NKAYDSHTAKSFELVNGYLTA-----SWVRPAgilRNGQTWEA--	265
Dd	497 GVDESKPYLDNEKULKVNSTLNLSGLTVNNTTNCKIQVGANGIKPATVANNVANTSATV	556
Qy	266 -----SNENDLRPVLMGWPPDKDTQAVYVMNMNYLSANETEVNTFNETSQVDLN	313
Dd	557 GTARITBEKIGFAGTND-----GVDBQAPFLD--KERLKVGRVEITFDGSG---IN	601

QY 1239 MDGNGNAVGOYPVNNOLLYFNPBTGVQVK--GQFITDAQRTSYDANGSALKSSGFFT 1296
Db 1500 ----LANAV-AQTP-----LTFAGDTGTAKKJAGETLTIKGGQT---DTNK-----LT 1538
QY 1297 PNGSDWYIABNGVYVYKGFQVAENQDQWYFYDQTTGQAKGAQKVDGRDLYFNPDSGVQV 1356
Db 1539 DNNIGVVAGTGTGTVKLAKDLT-----NLNSVAGGTIKDEKGISF-VDANGQA 1586
QY 1357 KGFDPATDESQNTSFYHG---DNGDKVVGFFTTGNNAWYADNNGNLVKGFOBID-GKWY 1412
Db 1587 K-----ANTPVSANGLDLGGKVI-----SNVGKGTQKTDAAVQ 1621
QY 1413 HFDVETGQQAAGAAVNGQOLYF-----DVIDSGIQ-----VKGDFYTDGCGTSSYVDVNS 1462
Db 1622 QLANEVRNLGLGNDNAGQVNIADIKDPPNSGSSNRRTVIKAGTVLGGKGNNDTEKLAT 1681
QY 1463 G-----DKKVGNGFFTTGD--NAW-----YYADGQ-----GNLAKGRKSI 1494
Db 1682 GGVOVGVDKQGN---ANGDLSNVVTKQDGSKKALLATYNAAGQTYVNTNPAEALDRI 1738
QY 1495 DNQDL-YFDPATQVKQGLVSDGRNYYFPDSSGN-----YADGQ-----MAKNRFVR 1537
Db 1739 NEQGIREFHVDNQGQBP----VVOGRNGIDSSASGKHSVAIGFQAKADGEAAVAIGRQIQ 1794
QY 1538 IGDWIIYFNGDGAAT 1552
Db 1795 AGNQSIAIGDNAQAT 1809

RESULT 8

US-11-013-759-11

; Sequence 11, Application US/11013759

; Publication No. US20050249747A1

; GENERAL INFORMATION:

; APPLICANT: Loomore, Sheena M.

; APPLICANT: Sasaki, Ken

; APPLICANT: Yang, Yan Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE

; TITLE OF INVENTION: PROTEIN OF MORAXELLA

; FILE REFERENCE: 1038-921MIS:jb

; CURRENT APPLICATION NUMBER: US/11/013,759

; CURRENT FILING DATE: 2004-12-16

; PRIOR APPLICATION NUMBER: US/09/361,619

; PRIOR FILING DATE: 1999-07-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 11

; LENGTH: 2314

; TYPE: PRT

; ORGANISM: Moraxella catarrhalis

US-11-013-759-11

Query Match 3.7%; Score 306.5; DB 7; Length 2314;

Best Local Similarity 21.0%; Pred. No. 2.6e-08;

Matches 353; Conservative 205; Mismatches 637; Indels 489; Gaps 91;

QY 23 SIGLSVLGAGTVSAEDKV-ANDTTAAQT-----VGVD-----TGQD---- 58

Db 776 SVGFVS-----TYNTVDFIDGNATTAKVTYDETNTQTSKYTYDVNVDEKTIETLGNGKTN 830

QY 59 -----QATTNDANTNTTDTADQSANNTQDQAGSDQSNQDQAKQDTANTDRNQADN 111

Db 831 KIGVKTTLTTLTNNAGKATNFSTNDALVN---AKDIAENLNTLAKE-IHTKGTADT 885

QY 112 S-----QTDNQATDQAT-SPATDGTSVQRDRDAANVATAADQSGTAPSQEKSAAALSDN 166

Db 886 ALQTFVKVKDGATDDTETIVGKDGTT--QNGKTVNTLKLKXENGLTVATNKGDTVTFGINT 943

QY 167 VK-LIPKYYVYVQAGSYKKNEFA----ITVNGQMLYFSDTGCALGSTSYFSQGTNLV 221

Db 944 QSLGKAGGSTTLNKDGLSISKNPASNEQIQVGADGVKFAKVDKGNSTGT-----DGTSRIT 999

QY 222 DD--FSSHNKAYDST-----AKSFELVNGYLTANSWYRPAGILRNGQTWASNE 268
Db 1000 KDQIGFTGANGSLDTTKPHLTQKLVGEVEITNTGINAGG--KKITNIQSGDITQNSND 1057
QY 269 -----NDLRPVLMSWFPDKDTQVAYV--NYNKKYLSANET-----EVTNETSQVDLAK 315
Db 1058 AVTGGVYDLKTELES-----KINSAAKTAQNSLHEFSVADEQCNHFTVSNPYSYDTSKT 1113
QY 316 AOSIQTKIBQKITSNDSTQWLRTAMEAFVAAQPKMNMSTENFNKGDLHGLGALLYTNSD- 374
Db 1114 SDVITFAGENGITTKVNGVVRGID-----QTK-GLTTPKLTVGN--NGKGVIVDSKDG 1166
QY 375 ----LTPWANSVRLNRTPTQDGTKKYFTGEGEGYFLLSNDVNSNPVVOAEQLNQL 431
Db 1167 QNTITGLSNT--LANVT--NDGAGHALSQ-----LANDTKTRAASIGDVLNAG 1212
QY 432 HYLNMWGDIVMGDKDANFDGVRVDAVDNVNADLLQVYSNYFKDNYKVTDESEALAHISI 491
Db 1213 FNLQNGEAV--DFVSTYD--TVDFIDG--NATTAKV-----TYDDTSKTSKVYDVNV 1260
QY 492 LEAWSLNDNQYNEDTNGTALSIDNSSRLTSLAVLTQKPGQRIDLSNLISSEVNKERANDT 551
Db 1261 -----DNKTIEVTSDEKLGKLVKTTT-----LTKTSA-----NGNATKFSRAD- 1296
QY 552 AYGDTIPTYSFVRAH-----DSEVQTVIAKIVKEKIDT---NSDGYTFTLDQLKDAFKIY 603
Db 1297 --GDALVKASDIATHLNTLAGDIQT--AKGASQASSASVVDADGNKVIVDSTDKKYQV 1352
QY 604 NEDMAKVNTYTHYNI PAAYALLLSNMESVPRVYGDLYTDDGQYMAKKSPPYDAIATML 663
Db 1353 N-DKQGVDK-----NKEVAKDKLVAQAQTPDG-----TLA 1381
QY 664 QGRI--AYVSGGSGSEVHKVNGNNOILLSSVRYGDLMSADDTQCTDLSTRSGLVTLVSNDP 722
Db 1382 QMNVKSVINKEQVNDANKQGINEDNAPFKIGLEN--AAKDTYKNAAVTVGDLNVAQTP 1439
QY 723 NLDLG-----GDSLTVNMGRAHANOAYR---PLILGTGKGVSQYLSKDSNTNIVKYT 770
Db 1440 LTFAGDTGTGTAKKLGELTITIKGGQDTNKLTDNNIGVWAGT--DGFTVKLAKDLTNL-NSV 1497
QY 771 DANGNLTPFADDIKGYSTVDMSGYLAVVVPVCAKQDQVRAADTNQKADGKSLKT-SAA 829
Db 1498 NAGG-----TRIDERGISFVDANGQAKANTPVLISANGDL-----GGRKISNIGAA 1543
QY 830 LDSQVIYEGFSNFQFANNADADYTNKKAENADFFKKLGITSFEMAPQVVSATDGSFLDS 889
Db 1544 VDD-----NDAVNFKQF--NEVAKTVNNLNQNSGASLPFVVTDANGKEPVDGPKQA 1597
QY 890 I--IQNGYAFSDRYDLAMSKNK--YGSKDDLANALKALHANGIQAIADWVPQIYQLPGE 946
Db 1598 IKGADGKYVHANANGVPVVDKQKPIITDADKLANL--AAHGKPLDA-----GH 1642
QY 947 EVVTAKRNTSYGNPTFDAYINNIALYATNTKS-----SGSDYQA 984
Db 1643 QVVASLGNS-----DAITLTNLIKSTLPQIDTPTNTGNANAGQASLPLSAAQ 1691
QY 985 QYGGAFDELKAKYPMFTVMISMISGKPIDPSTKIKQWEAKYFNGTNVNGKAGYVLSDD 1044
Db 1692 SNAASVKQVNLVNGFNLQTNHNOQVDFKAYD-----TVNFVNGT---GADITSVRSAD 1740
QY 1045 ATGKFTYVNEGDFLPASFTGDQNAKTGFYDGTGMAYYSTSCNKAIVNSFIYEGGHYVF 1104
Db 1741 GTMSNITVN-----TALAATDDDGNVLIKA----- 1765
QY 1105 DKDGHMVTGSKAEQNDYFPLNGIQMRDAIYQDAQ---GNSYYGRTGILYKGNWYVP 1161
Db 1766 -KDGKF--YKADD-----LMPNGLKAGKSASDAKPTGLSLVNPAGKSGTGD---- 1811
QY 1162 FVDPNNANKTVPRYFDANNVM-----AICGRNMYQTYFFDENGFGQAKQLLTD-DKGTH 1215
Db 1812 AVALNNLSKAVFKSKDGGTTTTTSSDGGISIQGKONSSITLSKDLGNLVNGKGVISNKGKTK 1871
QY 1216 VFDEDNAMAK--NKFVNVGDDWYMDGNGNAVKGQYPVNNQILYFNPBTGVQVKGQFIT 1273

```
Db 1872 DTDAAVQQLNEVRNLLGLGN-----AGNDNA-----DGNQVN---TA 1906
QY 1274 DAQGRSYDANSALKSSGFFTPNGSDWYIYAENGYYVYKGFQVAENQDQWYFYDOTTGK 1333
Db 1907 DIK-----KDPNSG-----SSNRTVIKAGTVLGG-----KGND-----TEK 1939
QY 1334 QAKGAAKVDGRDLVFNPDGSGVQVKGDPATDESNTSYFYHGDNGD-----KV 1379
Db 1940 LATG-----GVQV-----GVQKQDNA-----NGDLSNVVWKTQKQSKKA 1974
QY 1380 VGGFFTTGNNAIYDNNGLVKGFOEIDGKWHFDEVTCQQAALVNGQQLYFDVDS 1439
Db 1975 LIATYNAAGQTNLTNPAEADRINEQGIRFFHVND--GNQ---EPVVOGRN---GIDS 2026
QY 1440 -----J-----GIQVKGDFVTDGCGNTSYDVNSGDKKVGFFTTGDNAMYAD-----Q 1484
Db 2027 SASGKHSVALGFAK-----ADGEAAVA---IGRQTAGNQSTAIAGDQAATGDSIAIGT 2079
QY 1485 GNLAKGRKSIDNQDLYFDPATGKQVKGQLVSDGRNYYFDS-----GSGN---MAKNRF 1535
Db 2080 GNVVTGKHS-----GAIGDPSTVK--ADNSYSVGNNNQFIDATQDVFVGNNITVTESNS 2133
QY 1536 VRIG 1539
Db 2134 VALG 2137

RESULT 9
US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Shisong, Jiang M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; FILE REFERENCE: DFN-043CN
; CURRENT FILING DATE: 2004-12-22
; PRIOR FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-022-562-228

Query Match 3.7%; Score 302.5; DB 7; Length 396;
Best Local Similarity 24.8%; Pred. No. 4.1e-09;
Matches 108; Conservative 67; Mismatches 176; Indels 85; Gaps 18;

QY 1121 NDYY-----FLPNGIQMRDAIYQDAQNSYYIGRTGLYKGDNWWYFVDPNNANKTVF 1175
Db 1 NEYYPEIIVLPNTFHKVNIINLSSSFYKWSLEG-----SDFILVRV 44
QY 1176 FDANN---VMAIGYRNMVGYTYFYFDENGPOAK-----GQLTDDKGYHFDENGMA 1225
Db 45 LEBSNKKILQKIRIKGILSNTKSFNRMSIDFKDKLSLGYIMSNFKS---FNSNELDR 101
QY 1226 KN--KFNVNGDDWYDMGNGNAVGQYFVNNQILYFNPETGVQVKGQFITDAQ---GRTSY 1281
Db 102 DHLGFKIIDNKTYYDEASKLVKGLININSLFYFDP-----IESNLVTGWQTINGKKYY 156
QY 1282 YDANSALKSSGFFTPNGSDWYIYAENGYYVYKGFQVAENQDQWYFY--DOTTGKQAKGAA 1339
Db 157 FDINTGA-ASTSYKIINGKHFFYNNGVMQLG---VFKGPDGFEYPAPANTQNNIEGQA 212
QY 1340 KVDGRDLVFNPDGSGVQVKGDPATDESNTSYFYHGDNGDKVVGFFTTGNNAIYDNNGN 1399
```

```
Db 213 -----IVYQSKFLT---LNGKKYFNDNSKAVTGWQITDGKKYFNLNTAE 255
QY 1400 LVKQFEIDGKWHFDEVTCQQAALVNGQQLYFDVDSGIGVKGDFVTDGQGNNTSY- 1458
Db 256 AATGWQITDGKKYFNTNTSIASGTGVTIINGRHFYFNTDGMQIG---VFKGPNGEYFA 312
QY 1459 -----DVNSGDKKV-----NGFFTTGDNAMYADGQGNLAK---GRKSIDNQDLYFDPATGK 1507
Db 313 PANTDANNIEGOAIRYQNRFLYLDHNIYYF-----GNNSKAVTGWQTINGNVYFMPD 368
QY 1508 QVKGQLVSDGRNYYF 1523
Db 369 AAAGGLFEIDGVIIYFF 384

RESULT 10
US-11-052-554A-217
; Sequence 217, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 217
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-217

Query Match 3.5%; Score 290; DB 7; Length 726;
Best Local Similarity 22.4%; Pred. No. 4.1e-08;
Matches 142; Conservative 76; Mismatches 209; Indels 206; Gaps 28;

QY 946 EHVTAKTNSYGNPTFDAYIN---NALYATVTKSGSYQAYGGAFLDELKAKYPMDF 1002
Db 38 ETVITEQROTSKINASSQVENQTSNQVEA-KTDSANKDPQEKTSVATDA-----PSMN 91
QY 1003 TVNMISTGKPIDPSIKIQWEAKYFNGTNVLGKGAGYVLSDDATGKYFTVNENGDFLPAS 1062
Db 92 SANNMSQSD-----KQ-----NTVNEISSDSQQTDTBQTD-----LPQN 126
QY 1063 FTGDQNAKTGFYDGTGMAYYSTGNK-----AVNSFIYEG-GHYIYFDKGHMVTSYKA 1117
Db 127 SFKQQSAAHYKM-----TTEAKTSPSHSINTFVNDGNGWYVYLGADGRNVTGSH-- 174
QY 1118 EBGNDYFYLPNGIQMRDAIYQDAQNSYYIGRTGLYKGDNWWYFVDPNNANKTVF 1177
Db 175 ----- 174
QY 1178 ANNVMAIGYRNMVGYTYFYFDENGFOAKGQLLTDDKGT--HYFEDNGAMAKNFPVNVGDDW 1236
Db 175 -----TIG-----GKTMVFAQDGKQVKGAFQADSKNKHYYDRDSEGMTNFRVNDQGNW 224
QY 1237 TYMDGNGNAVKGYFVNNQILYFNPETGVQVKGQFITDAQGRTSYYDANSALKSSGFF 1296
Db 225 YYLNDGVPTGSIITVNGQSLYFNSD-GSQVKGNFV-BEDGSLRYDYDKNSGDLRLKRTSRT 282
QY 1297 PNGSDWYIYAENGYYVYKGFQVAENQDQWYFYDOTTGKQAKGAAKVDGRDLVFNPDGSGVQV 1356
Db 283 INGVNYQFDNDG-----NARAIDKIE-----VV 305
QY 1357 KGFATDESNTSYFYHGDNGDKVVGFFTTGNNAIYDNNGNLNVKGFQEBIDGKYH 1416
```

Db 306 KTSLVVD-----SYFSGPSVSKILFNFHVKVTPAVVHA-----GAMV-----T 343

Qy 1417 VTGQAK-----GAALVNGQOLYFDVDSGIGVKGDF-----VTDGQNTSYVDVNSG-----DKK 1466

Db 344 TAGVQRKILNSYVSNAGSHVYFVDSHSHVYVLELDIPYDPDNDSSRNASPFPIEDSNAFRNW 403

Qy 1467 VNGFPTGDNWYADQGN-----LAKGRKSIDNQDLYFDPATGK-QVKGQLVSDIGRNY 1522

Db 404 VNSYTVKVDNLQVADGNSQSIISSQDAINNR-----FLPTDTRFSRGRSGYGNFYAAYQ 460

Qy 1523 FDSGSGNMKNRFRVIGDQWY-FGNDGAATNL 1554

Db 461 PEAAIGG-EXNPLI-----VWLHGIGEVGTDINI 488

RESULT 11

US-11-052-554A-260

; Sequence 260, Application US/11052554A

; Publication No. US2005028866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: Patent in version 3.3

; SEQ ID NO 260

; LENGTH: 1647

; TYPE: PRT

; ORGANISM: Streptococcus pyogenes MGAS8232

US-11-052-554A-260

Query Match 3.5%; Score 286.5; DB 7; Length 1647;

Best Local Similarity 20.2%; Pred. No. 1.9e-07;

Matches 376; Conservative 227; Mismatches 650; Indels 611; Gaps 106;

Qy 1 MEKGLHYKLHKVKKHVVITIAVAGISGLVSLVGAGTVAEDKAVANDTTAQTATVGVDTGDDQA 60

Db 1 MEKKQRFSLRKYKSGTFSVLIGSVFLMW-----TTTVADELSTWSEPT 44

Qy 61 TTNDANTNTDITDADQSANTNQDQAGSQSNQDQAKQDPTANTDRNQADNQTNNQAT 120

Db 45 ITN-----HTQOQAQHLNTWLSAESKSD-TSQITPKTNR-----EKEQPQ 86

Qy 121 DQATSPATDGTSVORDDAANVATAADQEGTAPSEOEKSAALSLDNVKLIDGKYVYVQAD 180

Db 87 GLVSEPTT--TELAUTDAAPMA-----NTGPDATQKSASLPPVNTDVID--VWTKY 133

Qy 181 GSYYKNF-----AITVNG-----QMLYFSDTGTALSSTS----- 209

Db 134 GAWDKYKGGKGVAVVIDTGIDPAHQSMRISDVSTAKVSKEDMLARQAKAGINYGSWIN 193

Qy 210 -----TYSFQGTNLLVDD-FSSHNAVDSTAKSPFELVNGYLTANSWYRPAQ-----I 256

Db 194 DKVFAHNVENSNDIKENQFEDFEDWENFEDAEAPKAIKGHKIYRPSQSTQAPKETV 253

Qy 257 LRGOTWEASNDLRPVLMSWPKDQTVAYVN-----YNNKYL SAN--ETEVTNE----- 306

Db 254 IKTEET-DGSHDID-----WTQDIDDTKYESHGHHVTVGIVAGNSKEAAATGERFLGI 304

Qy 307 --TSQV-----DLNKEAQSIQTK-IEQKIT--SDNSTQWLR TAMEAFVAAQPKWNMS 353

Db 305 APEAQVMEFRVANDVWMSAESLFIKAIEDAVAGADVINLSLTANGAQLSGSKPLMEA 364

Qy 354 TENFNKGD-----HLOGGALLYTNSDLTPWA-NSDVRLL-----NRPPTQDQTKKYFTFG 403

Db 365 IEKAKKAGVSVVVAAGNVRVYSGDHDPLAIPNDYGLVGSPTGRTPT----- 412

Qy 404 GEGGYEFLLSNDVNSNPVQABQLNOLHVLNMGDIWVG-----DKDANFDGVRVDAVDN 459

Db 413 -----SVAAINSKWLIQ--RLMTVKELENRADLNHGKAIYSESVDFNKIYK--DS 457

Qy 460 VNADLLQVSNYSPKDN-----YKVTDSSEANALAHISILEAWSLNDNQYNEDTNGT---ALS 512

Db 458 LGYDKSHQFA-YVKESTDAGYKAQDVKK-----IALIE-----RDPNKTYDEMIA 502

Qy 513 IDNSSRLTSLAVLTQKQQRIDLSNLISSEVKNKERANDTAYGDTIP-----TYSFVRAHDS 568

Db 503 LAKKHGALGVLIFFNNKPGQ-----SNRSMRLTANGMGIPSAFISHEFGKAMS 550

Qy 569 -----EVQTVIAKIVKEKID-----TNSDGYTFTLDQLK-----DAFKI 602

Db 551 LANGNGTGSLEFDSVSVSKAPSKGNEMNHFNSWGLTSDGY-----LXPDITAPGDDIYST 604

Qy 603 YNEDMAKVNKTYTHYINIPAAAYALLLSNMESVPRVYGYDLYTDDQYMAKSPY--DAIA 660

Db 605 YNDN-----HYGSGTGTSM-----ASPQIAGASLLVK-QYLEKTQPNLPKEKIA-647

Qy 661 TMLQGRIAVSGGQSEVHKVNGNQLSSVR-YGQDLMSADDTQGTDLRTSGL-VTLV 718

Db 648 DIVKNLLM-----SNAQIH-VNPETKTTTSPROQAGALLNIDG-----AVTSLGYVTGK 695

Qy 719 SNDPNLDLGG--DSLTVNMGRAHANQAYRPLILGTGQVSYLKDSDTNIVKYTD---A 772

Db 696 DNYGSSISLGNITDTMTFDV-TVNN-----LSNKDKTLRY-----DTELL--TDHVDQ 740

Qy 773 NGNLTFADDIKGYSTVDMSGYLAVWVPVGAQGDQVRVAADTNQKADGKSLKTSAAlds 832

Db 741 KGRFTLTSLRSLKTYQGE-----VTVPANGK--VTVRVMTDYSQ-----PTKELTKQMSN 788

Qy 833 QVIYEGFSNPD-----PANNDADYTNKIKIAENADP-FKLGITSP---EMAP 876

Db 789 GYILEGFVRFRSDQDLNRVNI PFVFGQGFENLAVABESIYRLKSGQGTGFYFDSGP 848

Qy 877 Q---YVSA-----TDSGSL-----DSII 891

Db 849 KODIYVKGHFTGLVLTGSETNVSTKTI SNGHLTLTGTFKNADGKFILEKNAQNPVLAI 908

Qy 892 QNGYAFSDRYDLAMSKN---NKY-GSKODLANALKALHANGIOAIADWPDQIYQLPGE- 946

Db 909 PNG---DNNQDPAAPKGVFLRYQGLKASVYHASDKEHNPL-----WVSPESEFK--GDK 958

Qy 947 -----EVVTAKRNSVGNPTFDAYINNALYATNTKSSGSDYQAOYGGAFIDELKAKYPMDF 1002

Db 959 NFNSDIRFAKSTTLTGT-----AFSGKSLTGASLPDGYHYVWSY---YPDVVGAKRQEMT 1011

Qy 1003 TVNMISTGKPI-----DPSTKIKOWEAKYFNGTNVLGKA-----GYVLSDDAT 1046

Db 1012 FDMILDRQKPLVLSQATFDPTNRFPPEPLKDRGLAGRKDSVFLYERKONKPTVTINDS 1071

Qy 1047 GKYFTVNGDPLPASFTGD-----QNAKTG-FYDGTGMAYYSTSGNAKVSFIYEGGH 1100

Db 1072 YKYVSVEDNKT FVERQADGSFILPLDKAKLGDFFYMWEDFA-----GNVAIAKL---GDH 1123

Qy 1101 YYYFDKDHMTG-----SYKAECDNDYFPLPNGIQMRDAIYQAOQNSYYIGRTGLYKG 1156

Db 1124 L-----PQTLGKTPIKLTDGN--YQTKETLKNLEMTQSDTGLVNTQOALVAVHRN 1174

Qy 1157 D-----NWYPFVDPNNANKTVFRYDA--NNYMAIGRYNMVGYTYFPDENGFOA--- 1203

Db 1175 QPQSOLTKMNQDFFISPNEDGNKDFVAFKGLKNV-----YNDLTVNVYAKDDHQKQTPIW 1230

Qy 1204 ---KQQLTDDKTHYFDEDNAGMAKNFVNVDGDDWY---YMDGNGNAVKGQPVNNQIL 1257

Db 1231 SQAQASASAIESTAWY---GITARGSKVMFGDYQVYVTVYDEHCKEKKQYTTIS---- 1282

Qy 1258 YFNPETGVQVKGQFTIDAQGRTSYYDANSKALKSGFFTPNGSDWYABNGYYKGFQKV 1317

Db 1283 -----VNDKKPMIT--QGR---EDTINGV-----DHFTPDKTK-ALGSSGIVREBFVYL 1325

Db 1674 SIIFNGNDTIDPSKYQAGALIFASNGVSNINITLTNATNGLSNAGLNNSVQKGEICINL 1733
QY 1187 RNM-----YQOTYFDBNGFOAKGQLTDDK-----GTHYFDE 1219
Db 1734 ANCPFTTKNSPANSVPTPTNESLHVANNFTFLGTIISNGAIDLSQVNTNSVIGTILNE 1793
QY 1220 D-----NGAMAKNKFVNVDWYMDGNGNAVKQYPPVNNQILYFNPETGVQVKGQF--- 1271
Db 1794 NATLOANNLTITAFNNASNSTANIDGN-----FTLNOQATLSTNASGLNVMGNFSY 1846
QY 1272 -----ITDAQO-----RTSYIDANS- 1286
Db 1847 GDLVFNLSHVSRAINTQTATIMANNPLIOFNASSKEVGTYTILDSAKAIYGVNNO 1906
QY 1287 -----GALKSGSFTTPNGSDWYEAENGVYK-----GFKQVAENQDQWY 1325
Db 1907 ITGSSSLDNLKLYALIDINGKHMVMTDNLGLTYNGQAVSVKDGGLVVGFK---DSQNOYI 1963
QY 1326 YPDQTTGKQAKGAKVD-GRDLYFNPDSG-----VQVKGDFATDESNGTSFYHGDNGDK 1378
Db 1964 YTSILYNK-----VKIAVSNDPINNQPAPTLKQYIAIQGVQSVDSIDQAG---GNQAIN 2015
QY 1379 VVGFFFTTGN---AWYADNN-----GNLVKGFQFIDGKWHFEDEV-----T 1418
Db 2016 WLNKIFETKGSPLFAPYLYESHSTKDLTTIAGDIANTLEVIANPNFNDATNLIQNTYT 2075
QY 1419 GQOAKGA-----ALVNGQQLYFVDVSGIQVKGDFVTDG 1451
Db 2076 QQMSRLAKLSDTSTFARSDFLERLEALKNKRFAADAI PNAMDVILKYQSRRVKNVWATG 2135
QY 1452 QGNTSY-----YDVNSG-DKVNNGFTTGDNAWYADGQGNLAKGRKSIDNODLY-- 1500
Db 2136 VGGASPISGGTGLYGINVGYDRFKGVIVGGYAAVGYSGFHANITQSGSSNNVNVGVYSR 2195
QY 1501 -----FDBATGRQVQQLVSDIGRNNYFDS-----G 1526
Db 2196 AFIKSELWLSNETWGNKTFINSYDP-----LUSIINQSYRTDTWTDAKINYG 2246
QY 1527 SGNMAKNRFV---RIGDQWYFNGND 1549
Db 2247 YDFMFKDKSVIFKPKQVGLSYVYIGLSG 2273

RESULT 13

US-11-052-554A-228
; Sequence 228, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 228
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-228

Query Match 3.2%; Score 267; DB 7; Length 338;
Best Local Similarity 25.9%; Pred. No. 2.4e-07;
Matches 97; Conservative 53; Mismatches 126; Indels 98; Gaps 20;

QY 1145 YYYG-RTGILYKGNWTYFFVDPNNANKTVFYFDANNVMAIGYRNYMGYQTYTFENGFGQ 1202
Db 15 FFFGLLATNTVFANTTGGREFVDKNR-----KYYVKDDHKAIYMHKIDGKTYFFGDIGEM 69

QY 1203 AKGQLLTDDKGTHTYFDE--DNGAMAKNKFVNVDWYMDGNGNAVKQYPPVNNQIL--Y 1258
Db 70 VVCQVQLLEIPGTGYRNLFDNQPVNE--IGLQEKWYFFGQDGALLE--QTDKQVLEAK 123
QY 1259 FNPETGVQVKGQFITAQGRTSYTDANSALKSSGFTTPNGSDWYIYAEN----- 1307
Db 124 TSENTGKTVGEQYPLSAEKRTYTFD-NNVAVK-TGWIYEDG-NWYILNKLGNFGDDSYNP 180
QY 1308 ---GYVYKGFQ---VAENQDQ-----WYFDDOTGKQAKGAKVDGRDLYENPDSGVQV 1356
Db 181 LPGEVAKWTDQFHTIDIDRSKPAWPYILD-ASGRMLTDWQKVNKGWYTF----- 231
QY 1357 KGFDFATDESNGTSFYHGDNGDKVVGFFFTTGNNAWYADN-NGNLVKGFQFIDGKWHYFD 1415
Db 232 -----GSSGSMATGKYYVRGK--WYLDNKNKGDMKTGWQYLGKWKYYL- 272
QY 1416 EVTGOQAKGAALVNGQQLYFVDVSGIQVKGDFVTDGQNTSY- DVNSGDKKVNNGFTTG 1474
Db 273 ---RSSGAMVTGWYQ-----DGLTWYILNAGNGDMKTGWFOVNG 308
QY 1475 DNAWYVADGQGNLA 1488
Db 309 ---KWYIAYSSGALA 320

RESULT 14

US-10-793-626-1780
; Sequence 1780, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1780

Query Match 3.2%; Score 267; DB 6; Length 1155;
Best Local Similarity 19.3%; Pred. No. 1.2e-06;
Matches 294; Conservative 193; Mismatches 496; Indels 540; Gaps 78;

QY 1 MEKLLHYKLHKYKHHV---TIASVIGLVLGVAGTVAEDKVAANDTTAQTAVGVDITGQ 57
Db 9 INKRVDFLSNKNVKSIRKFTVTGTAI---LVGA----- 39
QY 58 DOATTNDANTNTTDTTADQSANVTQDQSGDSQNNQDQAKQDPTANTDRNQADNSQTDNN 117
Db 40 -----TLMFGAADNEAKAAEDNQLESASKEEQKGRSDNESSKLNQVD---LDNG 85
QY 118 QATDOATSPATDGTSVORRDAANVATAADQEQTPSEQEKSAALSLDNVKLDGKYYV 177
Db 86 SHSSEKTTNVNNAETVKVVEAPPTSDVSKPKANEAVVNTSEKTP----- 129
QY 178 QADGSYKKNFAITVNGQMLYFSDTGALSSTSYFSQGTTLNVDDFSSHKNKAYDSTAKS 237
Db 130 -----KITEAPTVEESI---AETPKSTTQ-----QDSTEKN 159
QY 238 FELVNGYLTANSWYRPAGILRNGQTWEASNENDLRPVLMSWVPDKDTQVAYVNNMKYLS 297
Db 160 -----NPSL-----KD-----N 166

Qy	298	ANETVNTNSQVD--LNKEAQSIQTKIEOKIETSDNSINTQWLRPTAMEAFVAAQPKWNMSTE	355
Db	167	LNSSSTTSKESKTDHSTKQAO--MSTNKSNLDTNDSPTQSEKTSQQA	212
Qy	356	NFNKGDHLQGGALLYTNSDLTPWANSDYRLNLTPTQODCTKKYFTEGGGGYEFLLSND	415
Db	213	-----NNDSTDNQSAFSPKQJDSKPSQKYTKTFNDEP-----TQD	248
Qy	416	VDNSNPYVQAEQLNQLHYLNMWGDIVMGDKDANFDGVRVDAVDNPNADL--LQVYSNYFK	473
Db	249	VEHTTTKLTPSIST-----DSSVNDKQ--DYTR-SAVASLGVDSETEAITNAVR	296
Qy	474	DNYKVTDSEANALAHISILEAW--SLNDNQYEDTNGTALSINDSSRLTSLAVLTQPGQ	531
Db	297	DNLDLKAASREQINEAIIAELKXDFSNPDYGDV--TPLAINTSQSKNS--PHKSASP	350
Qy	532	RIDLSNLSISVUNKERANDTAYGDTIPTYSFVRAHDSEVOTVIKIVKEKI-----	582
Db	351	RMNLSMLAABPNSGKNVNDKY-KITNPTLSLNKS--NNHANNVITPSTNEQFNULKANYELD	408
Qy	583	DTNSDGYTFTL-----DQLKD-----AFKIYNEDMAVKNKTYTHY--	617
Db	409	DSIKEGDTFIKYQYIRPGGLELPAIKQLRSKQSGIVANGYVDKTTNTTYTFNYVD	468
Qy	618	-----NIPAAAYALL-----LSNMESVP-----RVYGDLYTDDQYMAKSPYYDAI	659
Db	469	QYQNIITGSFDLIATPKRETAIKQNNQYPMVEITIANEVVKKDFIVDYN--KK-----DNT	521
Qy	660	ATMLQGRIAVYSGQSEVHKVGNQO-----ILSSVRYGO-----DLSMAD	701
Db	522	TTAAVANVDVNNKHNEVVY-LNQNNQNPKYAKYFSTVXNGKFIPEGVKYVEVTDTNAMV	580
Qy	702	DTQGTDLRSGLTVLNSPNDLGDGSLTVMNGRAHAN-----QAYRPLILG----	750
Db	581	DSFNPDLSNNSKVDVTSQFTPKVSADGTRVDINPARNWANGKKYIVTQAVRPTGTVNYT	640
Qy	751	-----TKDG-----VQSYLKSDPTNIVIKYTDANGNLFTTADDIKGYSTVDMSGYLAVWVPP--	801
Db	641	EYWLTRDGTNTNDFYRGTKSTTVTY--LNGSSTAQGDN--PTYSLGDY--VWLDKXK	692
Qy	802	-GAKDQDVRVAADTNOKADGSKLSKTSAAJLDSQVYIEGFSNQ-----DP-----	845
Db	693	NGVQDDDEKGLAGVYVTVLKDSNNRELQRTVTTDQSGHYQDFNLQNGTVTFVEFALPDNVTPS	752
Qy	846	-ANNADYTNKIAENADFFKGLGITSEFMAPOVVSATDGSFJLDSLIQNGVAFPSDRY--	901
Db	753	PANNS-----TNDADSDGE--RDGTRKVVVAKGTINNADNMTVDT--GFYLPKKNVG	802
Qy	902	DLAMSKNKYGSKDDLANALKALHANGIOAIADWVPDQIYQLPGEYVVTAKRTN--SYGN	959
Db	803	DYVWEDTNKGJQDD-----NEKGISNVK-----VTLKNXGDTIGT	839
Qy	960	PTPDAYTNALYATNWKSSSDYQAOYGGAPFLDELKAKYFDMFTVMNISTGKPIDPSTKI	1019
Db	840	TTTTD-4-SNGKYEFTGLENG-DYTIIEF-----ETPEGYTPTKQNSGSD-----	878
Qy	1020	KQWEAKYFNGTNVLKGAGVYVLSDDATGKYFTVNENGDFLPASFTGQDNAKTGFPYDGTG	1079
Db	879	---EGKOSNGTKTT-----VTVKDAUNK--TID-----SGYKFP--	907
Qy	1080	MAYYSTGNKAVNSFIYEGGHYYFYD--KDGHMVTVGSYKABEDGNDYIYFLPNGIOMRDALY	1137
Db	908	-----IYNLGDYVWEDTNKGJ-----IQDDSE-----KGISGVKVTIL	939
Qy	1138	QDAQNSYYVYGRGTGILYKGDWNWYFPFDPNNANKTVTFRYFDANNVMAJGY--RNMYGQTY	1194
Db	940	KDKNGNA-IGTTTTTDAASH--YQFKGLENGSYTV--EFETPS-----GYTPTKANSQDII	989
Qy	1195	YFDENGFOAKQOL-----LTDKKGTHYFDEEDNGAMAKNKFVNVCDDWYVWDGNGNAVKGO	1249
Db	990	TVDSNGIITTTGIIANGDNLITDSGFY-----KTPKYVSG-DYVWEDTNKGJQDD	1038
Qy	1250	YPVANNQILYFNPETGVQV-----KGQFI-----TDAQGRTSYDANSALKASSGFFTPNGS	1300

[illegible]

QY 1250 YPVNNQILYFNPETGVQV-----KGQFI-----TDAQGRTSYVDANSALKSSGFFTPNGS 1300

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 32.0752 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-38
Perfect score: 8237
Sequence: 1 MEKKLHYKLHKVKKHWITIA.....FVRIGDQWYFGNDGAATNL 1554

Scoring table: BLOSUM62
GapOp 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4014	48.7	1599	2 S22737	glucosyltransferas
2	3818	46.4	1518	2 A4811	glucosyltransferas
3	3635	44.1	1431	2 A45866	dextranucrase (EC
4	3472.5	42.2	1592	2 A38175	glucosyltransferas
5	3434.5	41.7	1475	2 B33135	gtfB protein precu
6	3429.5	41.6	1577	2 T30858	glucosyltransferas
7	3264	39.6	1449	2 T30857	glucosyltransferas
8	3243	39.4	1449	2 T30552	glucosyltransferas
9	3161.5	38.4	1375	2 J70345	dextranucrase (EC
10	3080.5	27.4	1365	2 A1483	glucosyltransferas
11	2964.5	36.0	1508	2 T31098	probable dextran
12	2799	34.0	1290	2 JCS473	dextranucrase (EC
13	654.5	7.9	2817	2 B97033	unchacterized pr
14	545.5	6.6	2710	2 A37052	toxin A - Clostrid
15	498.5	6.1	2364	2 I40884	cytotoxin L - Clos
16	464.5	5.6	2178	2 S5805	alpha-toxin - Clos
17	463	5.6	2366	2 S10317	toxin B - Clostrid
18	454	5.5	648	2 S10869	enterotoxin A - Cl
19	443.5	5.4	2367	2 S70172	toxin B - Clostrid
20	430.5	5.2	563	2 A37184	glucan-binding pro
21	357.5	4.3	2334	2 S32920	cell wall-associat
22	331.5	4.0	3229	2 A52221	dextranase inhibi
23	331.5	4.0	1829	2 E81086	iron-regulated pro
24	324.5	3.9	1829	2 S35027	cytotoxin RTX homo
25	323	3.9	2167	2 A71489	cell wall-associat
26	314.5	3.8	1463	2 T30290	AAS surface protei
27	298	3.6	1902	1 B44856	lactocepin (EC 3.4
28	285	3.5	1385	2 D89824	hypothetical prote
29	281.5	3.4	340	2 G95043	choline binding pr

30	280.5	3.4	3705	2 AD0123	probable autotrans
31	278	3.4	1315	2 T28679	fibrinogen-binding
32	277.5	3.4	2044	2 AB1180	probable peptidogl
33	276.5	3.4	721	2 C97980	endo-beta-N-acetyl
34	276.5	3.4	2399	2 H71879	toxin-like outer m
35	276	3.4	1305	2 H41662	150K mating aggreg
36	275	3.3	2529	2 B64635	toxin-like outer m
37	275	3.3	4936	2 AH2515	hypothetical prote
38	273.5	3.3	1904	2 T13256	tail-host specific
39	273	3.3	3844	2 T18402	asparagine/asparta
40	272	3.3	1441	2 B86807	hypothetical prote
41	270	3.3	1335	2 T30211	autolysin B - Stap
42	269.5	3.3	1821	2 AG2335	hypothetical prote
43	268.5	3.3	2348	2 AD1841	hypothetical prote
44	268.5	3.3	13055	2 T16580	hypothetical prote
45	267	3.2	338	2 A97914	choline-binding pr

ALIGNMENTS

RESULT 1

S22737
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: S22737; S28810; B44811; S22727
R:Jacques, N.
submitted to the EMBL Data Library, March 1992
A:Reference number: S22726
A:Accession: S22737
A:Molecule type: DNA
A:Residues: 1-1599 <JAC>
A:Cross-references: UNIPROT:Q00599; UNIPARC:UPI00000BEEF34; EMBL:Z11872; NID:g47530; PDB: 1A91
A:Experimental source: ATCC 25975
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase ge
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: S28810
A:Molecule type: DNA
A:Residues: 1-51 <GIF>
A:Cross-references: UNIPARC:UPI000017027C; EMBL:Z11873
C:Genetics:
A:Gene: gtfK
C:Keywords: glucosyltransferase; hexosyltransferase
F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match	48.7%	Score	4014;	DB	2;	Length	1599;
Best Local Similarity	51.7%	Pred. No.	1.5e-172;				
Matches	851;	Conservative	215;	Mismatches	430;	Indels	150;
Gaps	40;						
Qy	1	MEKKLHYKLHKVKKHWITIAVASIGLVSIVGAGTUSA-----EDKVANDTTAQATVGDV	54				
Db	1	MENKRYKHLKVKQWVTLAVASVALATIVGVSATSSLASAETNNNSPSTTVGEN	60				
Qy	55	TG-----QDOATTNDANT-NTTDTD-----TADQSANTN-QDOAGSDQSNQDOAQD	103				
Db	61	TNPVVEKVGTTTEVANTSNAITTEAETVADKPAAGTTVPNSGT-----T	106				
Qy	104	TDNRQA--DNSQTDNNQATDOATSPATDGTSVQRDAAANVATAA--DQEGQTAPSEKSA	160				
Db	107	SDRAAAVEAKEPETTAKPEVATKPTATTS--EVAANAGVAAPTTEKSKELSEAIEKA	163				
Qy	161	ALSLDNVKL-IDGKYYVQADGSKYKKNPAITVNGQMLYFSDGTGALSSTSTVSFGQTTN	219				
Db	164	AVSLDNIKKEDGKYLLLEDGSHKKNPAITVNGQVLYFD-ENGALSSTSTVSFTQETT	222				
Qy	220	LVDVDFSSHNKAYDSTAKSFELVNGVLTANSVVRPAGILRNQGTWEASNENDLRPVLSMW	279				
Db	223	LVTDFTKNNAAYDSTKASFELVDGYLTADSWYRPKBILEAGTTWKASTEKDFRPLLSMW	282				
Qy	280	PKDQTQVAYVNMNKKYLS-ANETE--VTNETSQVDLNKEAQSIQTKIEQKITSDSNTQWL	336				

Db 417 DPHMGDFPFLANDVDLSNPVQAEQNLQIHYLMWGSIVMGDKDANFGIRVADVNV 476
Qy 461 NADLQVSYNYFKDNYKVTDSANALAHISILEAWSLNDNQYNEDTNGTALSINDSSRLT 520
Db 477 DADMLQLYTNVFRYYGVNKSANALAHISVLEAWSLNDNHYNDKTDGALAMENKORLA 536
Qy 521 SLAVLTK-----QPQRDLNLSISVNERAND-----T 551
Db 537 LFLSLAKPIKERTPAVSPLYNNTFTTQRDEKTDWINKDGSKAYNEDGTVKQSTIGKYNE 596
Qy 552 AVGDTTPTYFVRAHDSVOTVIAKIVKEKIDTNSDGYTFTLQDLKDAFKIYEDMAKVN 611
Db 597 KYGDASGVYPIRAHDNVODIAEIIKKEINPKSGDFTTIDAEKQAFIYNKMLSSD 656
Qy 612 KTYTHYNIPAAAYALLSNMESVPRVYDLYTDDGQYMAKSPYDAIAITMLQCRATYVS 671
Db 657 KKYTLNIPAAAYAVMLQNMETITRVYVGDLYTDDGHYMETKSPYDITVNLMSRKIYVS 716
Qy 672 GQOSEVH-----KVGNNQILSSVRYGQDMLMSADDTQGTDLRSRTSLVTLV 718
Db 717 GGOAQRSYMLPTDGRKMDNSDVELYRTNEVYTSVRYGKDI MTANDTEGSKYSRTSGQVTLV 776
Qy 719 SNDPNLDLGGDS-LTVNMGRANQAVRPLILGTGQVQSYLXDS- - - - -TNIYKYTDANG 774
Db 777 ANPKNLQDSAKLANVEMGIHANQKTRALIVTAGDKINFTSDADAIAAGYKYKETDSNG 836
Qy 775 NLFTTADDIKGYSTVDMSGYLAVVWVPVGAKDQDVRVAADTNQADGK-SLKTSAAALDSQ 833
Db 837 VLTFGANDIKGYETPDMSGFVA VVWVPVASDNDQDIRVAPSTEAKKEGELTLKATEAYDSQ 896
Qy 834 VIYEGSFNODF--ANNDADYTNKKIAENADFPKLGITGSFEMAPQVVSATDGSFLDSII 891
Db 897 LIYEGSFNETQIPDGSPPSYNTRKIAENVDLFKSMGVTSFEMAPQFVSADDTGLDSVI 956
Qy 892 QNGYAFSDRYDLAWSKNNKYGSKDLANALKALHANGIOALADWPDQIYOLPGEVNTA 951
Db 957 QNGYAFADRYDLAWSKNNKYGSKDLRALKALHAGIOALADWPDQIYOLPGEVNTA 1016
Qy 952 KRTNSYGNPTFDAYINNALYATNTKSSGSDYQAYGCAFDELKAKYPMFTVMISTGK 1011
Db 1017 TRTDGAGRIADAIIDHSLYVANSKSGKDYQAKYGEFLAELKAKYPMFKVMISTGK 1076
Qy 1012 PIDPSTKI KQWEAKYFNGTNVLKGAGYVLSDDATGKYFTVNBENGDFLPASFTGDQNAKT 1071
Db 1077 PIDDSVKLQWKAEEYFNGTNVLERGVGYYVLSDEATGKYFTVTKEGNFIPLQLTGKEKVT 1136
Qy 1072 GFYDGTGMAYYSTGKAVNSFIYEGCHYVYFDKDHVMTGSKYKAEADGNDYY-FLPNCI 1130
Db 1137 GFSSDGKGYTFGTSQTAKSAFVTFGNTYYPDARGHMTNSEYSPNGKDVYRFLPNCI 1196
Qy 1131 QMRDAIYQDAQGNYSYVYGRGTLILYKGDNWPFFVDPNNANK-----TVFRYFDANNVMA 1183
Db 1197 MLSNIFYIDANGTYLYNSKGQYKGG--YTFKDVSETDKGSKVKVFRYFTNEGVA 1254
Qy 1184 IGRNNYMGOTYYFDENGFOAKQLLTDKRGTHYFDESDNGAMAKNFVNVYDGDYMDGNG 1243
Db 1255 KGVTVIDGTFQYFGEDGFOAKDLVTFKGTYYFDAHTGNGIKDWRNNGKRYWYFDANG 1314
Qy 1244 NAVKGQYFVNNQILFNPETGVQVKGQFTDAGRTSYSDANSALKSSGFFTPNGSDWY 1303
Db 1315 VAATGAQVINGOKLYFN-EDGSQVKGGVVKNADGTYTSKYKEGVELVTTNEFTTGDGNWY 1373
Qy 1304 YAENGYYVYKPKQVAENQDQWYFDDTTCQKQAKGVDRDLNFNPDGSGOVKGDPAFD 1363
Db 1374 YA-----GANGKTVTGAQVINGQHLNFAD-GSQVKGGVKN 1409
Qy 1364 ESGNTSFYHGDNDKVVGGFFTTGNNAWYADNNGNLVKGFOEIDGKWKYHFEVTCQAK 1423
Db 1410 ADGTYSKYNASTGERLTNEFFTTGDNWYIYGANGKSVTGEVKI----- 1453
Qy 1424 GAALVNGQOXYFDVDSGIVQKGFVTDGQNTSYDVNSGDKKXNGFFTTGDNAMYADG 1483
Db 1454 -----GDDTYPFKDGKQVKGQTVSAGNRSIYYGDSGKRAVSTWIEIQPGVYVYFDK 1507

Qy 1484 QG 1485
Db 1508 NG 1509

RESULT 3

A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866

R:Honda, O.; Kato, C.; Kuramitsu, H.K.

J. Gen. Microbiol. 136, 2099-2105, 1990

A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl

A:Reference number: A45866; MUID:91100958; PMID:2148600

A:Accession: A45866

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1431 <HON>

A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296

C:Keywords: glycosyltransferase; hexosyltransferase

F:1127-1146/Domain: cpl repeat homology <CP1>

F:1192-1211/Domain: cpl repeat homology <CP3>

F:1257-1276/Domain: cpl repeat homology <CP4>

F:1277-1297/Domain: cpl repeat homology <CP5>

F:1321-1340/Domain: cpl repeat homology <CP8>

F:1341-1361/Domain: cpl repeat homology <CP6>

F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 44.1%; Score 3635; DB 2; Length 1431;

Best Local Similarity 49.9%; Pred. No. 1.4e-155;

Matches 749; Conservative 218; Mismatches 422; Indels 112; Gaps 32;

Qy 1 MEKGLHYKLHKVKKHWTIIVASIGLSLVGAGTVAEDKAVANDTTAQTAVGVDTG- - - 57

Db 1 METKRRYKMKHKVKKHWTIVAVS-GLITL- - -GTTTGGSSVSAETEQOTSQSDKVVTKSDE 56

Qy 58 DOATNTDANTNTDITDQDSANTWQDQAGSQSNQDQAKQDT- - - - -ANTDR 106

Db 57 DKAAESSQTDAPKT- - - - -KQAQTEQTAQS-QANVADTSTITKSTPSONITTOANSDD 111

Qy 107 NQADNSQTDNNAQTAQATSPATDGTSVQRDDAANVATA-ADQEGQATPASEQEKSAAL-SL 164

Db 112 KTVTYTKSEAGTSEERTKQSEEAQTASSQALTQAKAELTKQRTPAQENKPNPVDLAAI 171

Qy 165 DNVLKIDGKYVYVQADSGYKKNFAITVNGQMLYFDSDTGALSSTSYFSQGTTLNLVDDF 224

Db 172 PNVKQIDGKYVYVIGSDGQPKKFNALTVNNKLVYFDKNTGALTDTTSQYQFKQGLTKLNDY 231

Qy 225 SSHNKAYDSTAKSFELVNGYLTANSWYRPAGLRNGQTWAEASNDLRPLVMSWPDKOT 284

Db 232 TPNHQIVNFENTSLTETIDNYVTADSWYRPKDILKNGKVTWASSSDLRPLMSWPDKOT 291

Qy 285 QVAYVYNNKYLANSATEVTNETSOVDLNLKEAQSIQTKLEQKITSNSTMWLTARTEAFV 344

Db 292 QIAYLNTNMQOGLGTGENVTADSSQESLNLAAQTQVVKIETKISQTOQTMURDIINSFV 351

Qy 345 AAQPKWNKSTE- - -NFKNGDHLQGGALLYTNSDLTPWANSYRLLNRTPTQDDGTKKYFT 401

Db 352 KQTPNWNQTESDTSAGEKDHQGGALLYSNSDKTAYANSYRLLNRTPTSGTKPKYPE 411

Qy 402 EGGEGGYFLSLNDVNDNPNVQAEQLNQLHYLMWGDIVMGDKDANFGVIRVADVNV 461

Db 412 DNSSGGYDFLLANDIDNSNPVQAEQLNQLHYLMWGSIVANDPEANFGVIRVADVNV 471

Qy 462 ADLLQVSNYKQNYKVTDSANALAHISILEAWSLNDNQYNEDTNGTALSINDSSRLTS 521

Db 472 ADLLQIASDYLKAHYGVDKSEKKNINHLISLEAWSNDPQYNKDKGAQLPINDKRLSL 531

Qy 522 LAVLTQPCQRIDLSN- - - - -LISESVNKEKANDTAYGDTIPTYSFVRAHDSVQ 571

Db 532 LVALTR- PLEK-DASNKNIRSGLEPVITNSLN-NRSEAGKNSERMANYIFIRAHDSVEQ 588
QY 572 TWIAKFKGKIDTNSGTYFTLDQKDAFKIYNEDMAKVNKYTHYNIIPAAVALLSNME 631
Db 589 TWIAKIIKAQINPKTDGLTFTLDQKDAFKIYNEDMRQAKKYTQSNIPTAYALMSNKD 648
QY 632 SVPRVYGDLYTDDGOYMAKKSPPYDAIATMLQGRITAYVSGGSEBHVKNVGNQ----- 686
Db 649 STRLYGMYSDGGQYMAKKSPPYDAITDLTKARIKYAAGQDMKITVEGDKSHMDWD 708
QY 687 ----ILSVRYGQDLMGADDTQGTDLRSRTSGLVTLVNSDPNLIDL-GDSLTVMGRAHANQ 742
Db 709 YTGVLTSVRYGTGANEATD-QGSEATKTQGMVITSSNPSLKLQNQDKVINVMGAHKKQ 767
QY 743 AYRPLILGTGQGVQSYLKSDT- NIVKYTDANGNLFTTADDIKGYSTVDMSGYLAWVPV 801
Db 768 EYRPLLLTTKDLGTSYTSDAAKSLYRKTRNDKGLVFDASDIQGYLNPQVSGYLAWVPV 827
QY 802 GAKGQDVRVAADTNQKADGKSLKTSAAALDSOVIVYEGFSNFQDPANNADADYTNKKIAENA 861
Db 828 GASDNQDVRVAASNKANATGVYESSALDSQLIYEGFSNFQDVTVKSDYTNKKIAQNV 887
QY 862 DFFKGLGITSFEMAPQVYSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANAL 921
Db 888 QLFKSGVTSFEMAPQVYSESDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGQQDMINAV 947
QY 922 KALHANGIOAIADWPQYIOLPGEVVTAKTNSYGNPTFFDAYINNALLYATNTKSSGSD 981
Db 948 KALHKSIGIOVIADWPQYIYNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGD 1007
QY 982 YQAOYGGAFDELKAKYPMFTVMISTGKPIDPSTKIKQWEAKYFNGTNVLGKGAGYVL 1041
Db 1008 YQAKYGGAFELSEAAKYPISFNRTQISNGKKIDPSEKITAWKAKYPNGTNILGRGVGYVL 1067
QY 1042 SDDATKGYFTVNEGDFLPASFTGQONAKTGFYDGTGMAYYSTSGNKAIVNFIYEG-GH 1100
Db 1068 KUNASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFTSTSGYQAQNSFVQDAKN 1126
QY 1101 YYYFDKDGHWVTGSYKAEQDNDYFIFLPNGIQMRDAIYDQAQNSYYTYGRTGILYKGDWNY 1160
Db 1127 WYYFDNNGHMVYGLQOL-NGEVQYFLSNGVQLRESPLENADGSKNYFGHLGNRY-SNGYY 1184
QY 1161 PFVDPVNANKTVPYREDANNMAIGYRNMYGQTYPYEDENGFOAKGQLLTDGK-THYPDE 1219
Db 1185 SF-----DNDSKRWYFDASGVMAVGLKTINGNTQYFDQGYQYKGAWITGSGDKKRYFD 1239
QY 1220 DNGAMAKNFVN-VGDDWYMDGNGNAVKGOYPVNNQILYFNPETGVQVKGFITDAQGR 1278
Db 1240 GSGMAVNRFANDKNGDWYLLNSDGLALGVQQTINGKTYFG-QDGKQIKGKIITD-NGK 1297
QY 1279 TSYDANSGALKSSGFFTPNGSDWYYAENGYYVKFKQVAENQDQWYYPDQTTGQAKGA 1338
Db 1298 LKYFLANSGLARNIPAT-----DSQNNWYFYG-SDGVAVTGS 1334
QY 1339 AKVDGEDLYFNPDSGVQVKGDFATDESGNTSFYHGDNGDKVCGFFPTGNNAWYADNNG 1398
Db 1335 QTIAKKLYFASD-GKQVKGSFVT-YNGKVKHYHADSGELQVNRFEADQGNWYLDNSG 1392
QY 1399 NLVKGFQETDGKWHYHDFEVTGQAKGAALVNGQQLYFVDVDSGIVQKGFVFTGQGNNTSY 1458
Db 1393 EALTGSQRI-----NDQRVFF-TREGQVKQGVADYBERGLLRY 1430
QY 1459 D 1459
Db 1431 D 1431

RESULT 4
A38175
glucosyltransferase precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004
C:Accession: A38175

R;Abo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A:Reference number: A38175; MUID:91123227; PMID:1704006
A:Accession: A38175
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1592 <ABO>
A:Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:g217032; PIDN:BAAL4241.1; PID
F:1093-1112/Domain: cpl repeat homology <CP1>
F:1227-1241/Domain: cpl repeat homology <CP2>
F:1287-1306/Domain: cpl repeat homology <CP3>
F:1330-1351/Domain: cpl repeat homology <CP4>
F:1352-1371/Domain: cpl repeat homology <CP5>
F:1402-1420/Domain: cpl repeat homology <CP6>
F:1465-1484/Domain: cpl repeat homology <CP7>
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 42.2%; Score 3472.5; DB 2; Length 1592;
Best Local Similarity 45.1%; Pred. No. 3.3e-148; Mismatches 496; Indels 153; Gaps 39;
Matches 741; Conservative 254;
QY 1 MEKLLHYKLHKYKHWVTIAVASIG-LVSLVGAGTVSAEDKVVANDTTAQATYGVDTGDOQ 59
Db 1 MEKNVRFKMHKVKKHWVTLVASATMLASALGASVASADTDTASDDSNQTVV---TG-DQ 56
QY 60 ATTNDQ--NTNTTDTTADQANTNQDQAGSDQSNNOQAKQDPTANTDRNQADNSQTDNN 117
Db 57 TTNNQATDQTSIAATATSEQASATD---AATDQASAAEQTCGTASTD--TAAQTITNAN 111
QY 118 QATDQATSPATDG-TSVQRDAANVATAADQEGQTAPSEQEKSAALSLDNKLIDGKYYY 176
Db 112 EAKWVPTENQGFDEMLAEAKNVATA---ESDSIPSDLAK-----MSNVKQVDGKYYY 163
QY 177 VOADGSKYKNFAITVNGQMLYFSDTSGALSSSTYSFSQGTITNLVDD---FSSHKNAYDS 233
Db 164 YDQDGNVKNFAVSGDKIYYFD-ETGAYKDTISKVDADKSSSAVSONATI PAANNRAYST 222
QY 234 TAKSPELVNGYLTANSWYRYPAGILRNGQTWASENDLRPVLMSWPKDQTOVAVYNNY 293
Db 223 SAKNFEADVNYLTADSWYRPKSILKDGKTWTESGKDDFRPLLMWWPDTETKRYNVYNN 282
QY 294 KYLSANEYEVNETSQVDLNKEAQSIOTKIEQKITSDNSTOWLRTAMAEFAVAQAQKNMS 353
Db 283 KVVGIDKT-YTAETSQADLTAAAEVLQARIEQITSENNTKWLREAISAFVTQPMNGE 341
QY 354 TENFNKGDLHLOGGALLYTN-SDLTPWANSYRLLNRTPTQODGTK---KYFTEGEGGY 408
Db 342 SEK-PYDDHLQNGALLFDNQDTLTPTQSNYELLNRTPTNQTSGLDSLDSRFTYNPNPLGGY 400
QY 409 EFLLSNDVDSNPVVQASQNLNOLHYLMWGDIVMGDKQANFGVRVDVADVNNVADLLQVY 468
Db 401 DFLLANDVDSNPVVQAEQLANWLHYLLNFGSIYANDADANFDSIRVDAEDNVDADQLQIS 460
QY 469 SNYFKDNYKVTDSANALAHISILEAWSINDQNYNEDTNGTALSIDNSSRLTSLAVLTQK 528
Db 461 SDYLKAAAYGIDKKNANNHVSIVEAWSNDTTPYLHDDGDNLMNMDNKFRLSLMLSLAKP 520
QY 529 PQCRIDLNLNLSIESYKNERANDTAYGDTIPTYSFVRAHDSSEVQTVIAKIVKSKIDTNSDG 588
Db 521 TDVRSGPLIHNLSL-VDREVDREVEIVPSVSFARAHDSVEQDIIRDIIKAEINPNSFG 579
QY 589 YFTFTLDQLKADPKIYNEDMAKVNKYTHYNIIPAAVALLSNMESVPRVYTGILYTDGQY 648
Db 580 YSFTQEEIDQAPKIYNEDLKKSDDKKYTHYNVPLSYTLTLTNKGSIPRVYTGDMFTDDGOY 639
QY 649 WAKKSPYYDATATMLQGRITAYVSGGSEBHVKNVGNQILSVRYGQDILMSADDTQGTDL 708
Db 640 MANKTVMYDAIESLLKARKMYVAGQAMQNYOI-NGEILTYSRVYKGGKAKQSD-KGDAT 697
QY 709 SRTSGLVTLVSNDPNLDLGGDSLTVNMGRAHANQAYRPLILGTGQGVQSYLKSDTNN--- 765
Db 698 TITSGVGVVMGNQPNFSLDGKVVVALNMGAAHANQEYRALMVSTKDG VATYATDADASKAG 757

Db	688	GKALNATDT-GDRTTRTSQVAVIEGNNPSLRKASDRVVVNMGAHKNQAYRPLLLTTD	746
QY	753	DGVQSYLKSD--TNIVKYTDANGNLFTTADDIKGYSTVDMSGYLAWVPVGAK--DGQDV	809
Db	747	NGIKAYHSDQEAAGLVRYTNDRGELFTTAAIDKGVANPQVSGYLVGVVPVGAALIKMFAL	806
QY	810	RVAADTNQADGKSLKTSAAALDSOVIYEGFSNFDQFANNADYTNNKKAENADFFKKLGI	869
Db	807	RLARPHQQA--SVHQNALDSRVNFEFSNPAFAATKEEYTNVVIARNVDKFAEWGV	863
QY	870	TSEMAPQVVSATDGSFLDSIIQNGYAFSDRDYDLAMSKNNKYGSKSDLANALKALHANGI	929
Db	864	TDFEAPQVVSSTGSGFLDSVLQNGYAFTRDYLGISLKNKYGTADDDLKAIKALHSKI	923
QY	930	QAIADWPQIYQLPGEVVTAKRTNSYGNPTFDAYINNALYATNTKSSGSDYQAYGGA	989
Db	924	KVMADWPQWYAFPEKEVVTATRVVDKYGTPVAGSQIKNTLYVVDGSKSGDKQAYGGA	983
QY	990	ELDELKAKYPDMFTVMIMSTGKPIDPSTKIKOWEAKYFNGTAVLKGAGYVLSDDATGKY	1049
Db	984	FLEELQAKYPELFARKQISTGVPMDSVKIKOMSAKYFNGTNILGRGAGYVLKDAQNTY	1043
QY	1050	FTVNEGD--FLPASFTGQNAKTFYDGTGMAYYSTSCNAKAVNSFIYEGGHYYFDRD	1107
Db	1044	FNISDNKEINFLPKTLL--NQDSQVGSYDGKGVYYSTSGYQAKWTFISEGDKWYFDDN	1102
QY	1108	GHMVTGSYKAEDGNYYFLPNGIQMRDAIYQDAQNSYYVYGRGILYKGDNNYPPVDPN	1167
Db	1103	GYMVTGA--QSIGNVYVYFLSNGLQLRADAILKNEIDGTYYVYDNGRRYE--NGYQFM	1156
QY	1168	ANKTPRYPDANNVMAIGYRNMYGTYIYDENGFQAKGQLLTTDDKG--THYFDEDNGAMAK	1226
Db	1157	--SGVWRHFN--NGEMSGLTVIDGQVQYFDEMGYQAKGFVTTADGKIRYFDKQSGNMYR	1213
QY	1227	NKPV--NVGDWYMDGNGNAVKGOYVNNQILYFNPEGTGVQVKGFITDAQRTSYDAN	1285
Db	1214	NRFIENEKGWLYLGEDGAATGSGTQINGQLYFR--ANGVQVKGEFVTDHGRISYDGN	1272
QY	1286	SGALKSGSFTPTNGSDMYAENGYYVYKPKQVAENQDQWYFDTTQTKQAKGAAKVDRD	1345
Db	1273	SGDQIRNR-----VRNAQGWYFED--NNGYAVTGARTINGQL	1309
QY	1346	LYENPDSGVQVGDPRATDESNTSPYHGDNGDKVVGGFPTTGNNAWYADNNGNLVKGFQ	1405
Db	1310	LYFRA--NGVQVKGEFVTDYGRISYYDNGSGDQIRNFRVRNAQGWYFDDNGYAV----	1364
QY	1406	EIDGKWYHDEVTGQAKGAALVNGOOLYFVDVDSGIVQVKGDFVTDGQGNYSYDVNSGDK	1465
Db	1365	-----TGARTINGQLYFRA--NGVQVKGEFVTDHGRISYYDNGSGDQ	1406
QY	1466	KVNGFPTTGDNAWYADGGGNLAKRKSIDNQDLYPDATGKQVKQLYSID--GRNYYPD	1524
Db	1407	IRNFRVRNAQGWYFDDNGYAVTGARTINGQLYF--RANGVQVKGEFVTDYGRISYYD	1465
QY	1525	SGSGNMAK 1532	
Db	1466	ANSGERVR 1473	

RESULT 6

T30858
Glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30858
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: Z20909; PMID:95122197; PMID:7822030
A:Accession: T30858
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-1577 <SIM>

Db	951	PSOYTRVRAQNAKLFKEMGITSFEPAPQYVSQDGTFLDSIIENGAFEDRYDIAMSKN	1010
Qy	909	NKYGSDDDLALNALKALHANGIOAIAADWPDIQYQLPGEVVTAKRTNSYGNPTFFDAYINN	968
Db	1011	NKYGSLKDLMDALRALHAEGISAIAADWPDIQYINLPGEVVTASRTNSYGTFRPNAEYIN	1070
Qy	969	ALYATNWKSGSYQAOYGAFIDELKAKYPDMFTVMNISTGKPIDPSTKIKOWEAKYFN	1028
Db	1071	SLYAAKTRTFGNDFOGQYGGAFIDELKAKYPAIFERVOISNGRKLATTNEKITQWSAKYFN	1130
Qy	1029	GTNVLCKGAYLSDDATGKYFTVNEGDFLPASFTGDQNAKTFPYDGTGMAYYSTGN	1088
Db	1131	GSNIQGTGARYLQDNATNQYFSVKAQOTFLPKQMT--EITGSGFRVRGDDVQYLSIGGY	1188
Qy	1089	KAVNSIYEGGH--YYFDKDGHMVTGSYKAEDGNDYFPLNGIQMDALYQDAQGSYYY	1147
Db	1189	LAKNTFIOVGANQWYFDRKGNWVTGE-QVIDGKYFFLDNGLQLRHVLRQSGDHVYYY	1247
Qy	1148	GRGTI-LYKGDNMWYFPVDPNNANKTVFRYFDANNVMAIGYRNMNYGOTYYFDE-NGFQAKG	1205
Db	1248	DPKGVQAFNG--FYDFAGP---RODVRYPDNGQMYRGLHDMYGTTFYFDEKGTGIAQKD	1301
Qy	1206	QLLT-DDKGTHTFDEPDNGAMAKNFVNGDD--WYMDGNGNAVKQYPPNNQILLYFNPE	1262
Db	1302	KTRFADGRTRYPIDPTGNLAVNRFAQNPENKAWYYLDSNGYAVTGLQTINGKQYFYDNE	1361
Qy	1263	TGVQVKGQFITDAQGRTSYVDANSALKSSGFRPTNGSDWYYAENGYYVYKGFQVAENQD	1322
Db	1362	-GRQVKGHVT-INNRQYFLDGSGBIAFSRFTEN-----N	1396
Qy	1323	QWYFDQTTGKAQGAAGVDRGLYFNPDSGVQVKGDFATDESGNTSFYHGDNGKRVVGG	1382
Db	1397	KWYVVD-GNGKLVKGAQVINGNHYYFNNDYS-QVKGAWA-----NGRYYDGSQGAVSQ	1449
Qy	1383	FTTGNWAMYADNNGNLVKGQEIQDKYHDEVTGQQAAGALVNGQQLYFDVDSGTQ	1442
Db	1450	FIQIAANOWAYLNQDGHKVTGLQINNNKYIFG-----SNGAQ	1487
Qy	1443	VKGDFVTDGQNTSYVDVNSGDKVNGFFTTGDNAMYADGQNLAKGRKSIDNQDLYPD	1502
Db	1488	VKGKLLT-VQGKCFDAHTGQVNVRFVEARGCWYTFNSAGQAVTGGQVNGKQLYPD	1546
Qy	1503	PATGKQVKGQVLSIDGRNTYFDSGSGNMAKNR	1534
Db	1547	-GSGRQVKGRYVVGSKRFLFCDAKTGELQRR	1577
RESULT 7			
T30857			
Glucosyltransferase - Streptococcus salivarius			
C.Species: Streptococcus salivarius			
C.Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004			
C.Accession: T30857			
R.Simpson, C.L.; Giffard, P.M.; Jacques, N.A.			
Infect. Immun. 63, 609-621, 1995			
A.Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri			
A.Reference number: Z20909; MUID:95122197; PMID:7822030			
A.Accession: T30857			
A.Status: preliminary; translated from GB/EMBL/DBJ			
A.Molecule type: DNA			
A.Residues: 1-1449 <STM>			
A.Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166B; EMBL:L35495; NID:g662378; PID			
C.Genetics:			
A.Gene: gtfL			
Query Match			
Best Local Similarity 44.3%; Pred. No. 6.7e-139;			
Matches 729; Conservative 197; Mismatches 423; Indels 296; Gaps 47;			
Qy	1	MEKKLYHLKHVKHWTIAVA--SIGLVSUVAGT-----VSAED-----	39
Db	1	MDKKVHYKMHVKQWVTIATVGLSLGAVSAVSLGTGVDGVQADEHTATVAIPDITVDT	60

Qy	40	-KVANDTTA--QATGV-----DTGQDQAT---TNDANTNTTDT---DTADQSANTNODQA	86
Db	61	GTVSNDDTTAAQPTTAAVAATNDVATDQATPTATFOLDTTDTTVAANAADVTATVATGTDRA	120
Qy	87	GSQDSNQQAQKQDPTANTDRNQADNSQTNQNAQATDQATSPATDGTGTSVQR-----	135
Db	121	A---TTNDTTATNDTATDVTTN--NNTTTTDTTVDRAATTERRATGARRGTGGRATPV	175
Qy	136	---RDAANVAT-----AADQEGQAPSEQEKSAAALSIDNVKLDGKYYVQADGSYK	184
Db	176	NGNTNNWNTVTVVNDLPAATNNVVDGFS-----HIKTINGKQYVYVEDDGTIR	224
Qy	185	KNFALTIVNGQMLYFSDDTGALSSTSYSPSQ--GTTNLVDDFSSH-----NKAYDSTA	235
Db	225	KNVVLERIGSGYFNAETGELSNOKEYRFDKNGGTGSSADSTNTVTVNGDKNAFYGTTD	284
Qy	236	KSFELVNGYLTANSYTRPAGILRNGQTWEASNENDLRPVLMSWPKQKQVAVYNNKY	295
Db	285	KDIELVDGYFTANTWYRPEIKLDGKEMTASTENDKRPLLTVMWPSKAIQASYLNTMKEQ	344
Qy	296	-LSANETEVTNETSQVDLANKAQSLOTKLEQKITSNSTMQLRTAMEAFVAAQPKNMST	354
Db	345	GLGTQGT-YTSPSSQTDQDAALEVQKRIEERIAREGNTDMLRTTIKNFVKTPQGNST	403
Qy	355	ENFNKGDHLQGGALLYTNSDLTPWANSDYRLNLRPTQODG--TKKYFTGEGGEGYEFLL	412
Db	404	ENLNDNDHLQGGALLYNNDSRTSHANSDYRLNLRPTSTQKHNPKYTKDTSNGGEFLL	463
Qy	413	SNVDNSNPVQAEQLNQLHYLMNMGDIVMGDKANFDGVRVDADVNDVNADLLQVYSNYF	472
Db	464	ANDIDNSNPAVQAEQLNQLHYLMNMGDIVMGDKANFDGVRVDADVNDVNADLLQVYSNYF	523
Qy	473	KDNYKVTSEANALAHISILEAWSLNDNQYNDTGTALSIDNSSRLTSLAVLTQPGQR	532
Db	524	KAKYGADQSQDAIKHLSILEAWSLNDNQYNDTGTALSIDNSSRLTSLAVLTQPGQR	593
Qy	533	IDLNLISVSNKERANDTAYGDTIPTYSFVRAHDSVQTVIAKIVKEIDTNSDGYTPT	592
Db	584	SGVEPLISLNSL-DRSESGSKSRMANYAFVRAHDSVQTVIAKIVKEIDTNSDGYTPT	642
Qy	593	LDQLKDAFKIYNEDMAKNVYTHYNIIPAAYALLSNMESVPRVYVYDLYTDGQYMAKK	652
Db	643	LDEMKKAFIYNKDMRSANKQYQYNIIPSAVALMTHKDTVPRVYVYDLYTDGQYMAKK	702
Qy	653	SPYDAIATMLOGRYAYVSGGSEEVHKVN-CNNQ-----ILSSVRYQDLMSADDTQ	705
Db	703	SPYDAIETLLKGRIRYAGGGQDMKNYIGYNTNGWDAAAGVLTSVRYGTGANSASDT-G	761
Qy	706	TDLRSRTGLVTLVSNDPNLDLGGDSLTVNMGRAHQAQYRPLILGTQGVQSYLKDSDTN	765
Db	762	TAEATRNQMAVIVSNQPALRL-TSNLTINMGAAHNRQAQYRPLLLTNDGVATYVINDSDAN	820
Qy	766	-IVKYTDANGNLFTTADDIKGYSTVDMGGYLAWVPVVGAKDQDQVRAADATNQKADGKSL	824
Db	821	GIVKYTDGNGNLTFSANBIRGNPQVGVYLAWVPVVGASENQDQVRAVPSKEKNSSGLVY	880
Qy	825	KTSALDSQVIYEGFSNQDPAANDADYTNKIAENADFFKLGITSPMAQVYSATDG	884
Db	881	ESNAALDSQVIYEGFSNQDPAANDADYTNKIAENADFFKLGITSPMAQVYSATDG	940
Qy	885	SFLDSIIQNGYAFSDRYDLAMSNNKKNYKSDDLAALKAHANGIOAIAADWYPDQYQLP	944
Db	941	SFLDSVIQNGYAFSDRYDI GSKNNKYGLSLADLKAALSHAVGISAIDWYPDQYQLNLP	1000
Qy	945	GEEVVTAKRTNSYGNPTFDAYINNALYATNTKSSGSDYQAOYGGAFDELKAKYPMFTV	1004
Db	1001	GDEVVTVATRVNNGYGETKGAIIIDHSLYAAKTRTFGNDYQYGGAFDELKELYPQIFDR	1060
Qy	1005	NMISTGKPIDPSTKIKOWEAKYFNGTIVLKGAGVYVLSDDATGKYFTVNEGDFLP----	1060
Db	1061	VQISTGKMTTDEKITQWSAKTMNGNTILDRGSEYVLKNGLNG-YYGTINGGKVSPLKPVVG	1119

Qy 1156 -GDNWVYFDPNNANKTVFYFDANNVMAIGYRNMVGYTYFFDENGFAKGLLTDGK- 1213
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
1185 FGQW-----RYFQ-NGTAVGLTRVHGAVQYFDASGFQAKGQFITTTADGK 1229

Qy 1214 THYFEDNAGAMAKNFV-NVCDHWYVDGNNAVKGOYPVNNOILLYENPETYGVQVQFI 1272
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
1230 LRYFDRSGNQISNRFRVNSKGFWFLPDHNGVAVTGTTFNGORLYFKP-NGVOAKGEFI 1288

Qy 1273 TDAQRTSYDYDANGSALKSGFFTPNGSDWY-YAENGIVYVGKFQV 1317
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
1289 RDANGYLRYDPSNGNEVRNRFVNSKGFWFLPDHNGIATGARVV 1334

RESULT 10
A14183
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus'
C:Species: Streptococcus sobrinus
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Mar-2004
C:Accession: A41483
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
Infct. Immun. 58, 2452-2458, 1990
A>Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltransferase
A:Reference number: A41483; MUID:90316665; PMID:2142479
A:Accession: A41483
A:Molecule type: DNA
A:Residues: 1-1365 <GIL>
A:Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:
C:Genetics:
A:Gene: gtfS
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 37.4%; Score 3080.5; DB 2; Length 1365;
Best Local Similarity 43.9%; Pred. No. 1.1e-130;
Matches 667; Conservative 213; Mismatches 446; Indels 193; Gaps 34;

Qy 1 MEKKLHYKLHKVKGHWITIAVASGLVSLVGAGTVSAEDKVANDTTAAQTGVDVTGDQA 60
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
1 MEKNRLYKLHKVKQWVAIGTVTTLTSLAGGQVAA-----DTNNNDG 44

Qy 61 TTNDANTNTTDDTADQSANTNQ--DQAGSQDSNNQDAKQDPTANTDRNADNSOTDN--- 116
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
45 TSQVQNKVPSPDKFDAQAQNGQLAQMFAANQADQ----TATSQVSPTDGRVDNQVT 100

Qy 117 -----NOATDQATSPATDGTSVORRDAAVATAAQEGOTAPSEQEKSAALSLDNVK 168
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
101 PAANQPAANVANQDVANPATDAGLNKFSRAADTSF----DGKAVPQTSDPG-----HLE 151

Qy 169 LIDGKYVYVQADGSYKKNKPAITVNGQMLYFPDSDTGALSSSTYSFSQGTNLVDDFSHN 228
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
152 TVDGKTYVVDANGQRLLKNYSMVIDGKTYVFDGGTGEAQTDLPKTGQANQDNVPDSYQANN 211

Qy 229 KAYDSTAKSFELVNYGLTANSWYRPAGLRNGQTWEASNENDLRPLVMWWPKDQTOVAY 288
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
212 QAYSNEASSFTVDNYLTADSWSYRPRILKXGOSQWASSEGDLRPIILTWMPDAATKAAY 271

Qy 289 VNTMKNKYLASNETVETNETSQVDLNKEA--OSIQTKIEOKITSDNSQWLRTAMEAFVAA 346
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
272 ANFW-----AKEGLISGSYQNLSANLDAATQNIQSALEKKIASEGTNWLIRDMSQVFKS 326

Qy 347 QPKWNMNSTEN---FNKGDLHQGGALLYTNSDLTPWANSYRLNRTPTQQDGTKKYFTEG 403
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
327 QNQWSIASENETVYPNQDHMGCGALLFSNSKOTEHANSWRLLNRTFTFGTKQKYFTT- 385

Qy 404 GEGGYEFLSDVNSNPVQAEOLNQHLYLMNMGDI VMGDKOANFCGVRVADVANNVAD 463
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
386 NYAGYEILLANDVNSNPVQAEQLNHLHLYLMNMGDI VMGDKOANFCGVRVADVANNVAD 445

Qy 464 LLQVYSNYFKDNYKVTDSEANALAHISILEAWSLNDNOYNEDTNGTALSIDNSSRLTSLA 523
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
446 LLQIQDYIKAKYGTDOQNEKNADHLHSILEAWSGNDNDYVKDQNFSLSIDNDQRSGLK 505

Qy 524 VLTKQPQQRIDLNLISSEVNSKERANDTAQDITIPTYTSFVRAHDSEVQTVIAKIVEKI - 582
Db ||| :||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Qy	1113	GSYKAEDGNDYYFLP	NGIQMRDALYQDAQGSYYVYGRGTILYKGDNNYFPVDPNNANKTV	1172
Db	942	GN-QVINGINYYFLP	NGIEIQDA-YLVHDSGMYIYNNIG-KQLHNTYYQDKQRK	992
Qy	1173	FRYFDANNVMAIGVRNNYQ	-----TYIFDENGFOAKQQLLTDDKG-THYFDEDNAMAK	1226
Db	993	FHYFFEDGHWAQGI	VTIIQSDGTPVTQYFDBENGKQKQGVAVKSGDSGHLHYFDGASGNMLF	1052
Qy	1227	NKFPNVGD-DWYYMDGNGNNAVK	QYPVNNQILYFNPETGVQVKGQFITDAQGRTSYYDAN	1285
Db	1053	KSMGRADGSLVYDEKGN	AVTGQTINNTVYFN-DGGRQIKNNFKELADG	1103
Qy	1286	SGALKSSGFTPN	SGSDWYYAENGYYVYKGFQVAENQDQWYFYDQTTCKQAKGAQVGDGR	1345
Db	1104	-----SMVLYLN	-----KGVA-----VTGEQI-----INGQT	1125
Qy	1346	LYFNPDSGVQVKGDFATDES	GNSTFYHGDNGDKVVGFFTTGNNAWYADNNGNLVKGQF	1405
Db	1126	LYFGND-GRQFKGTTHINAT	GESRYYPDPSGNMITDRFERVGDNQWAY	1172
Qy	1406	EIDGKYHFEVTVQQAAG	AALVNGQOLYFVDSGIVKVGDFVTDGQGNSTSYDVNSGDK	1465
Db	1173	-----FGYD-----	GVAVTGDRIIKGQQLYFN-QNGIQMKGHLRL-NGIMRYYDADTGE	1221
Qy	1466	KVNGFFTTGNNAWYADG	QGNLAKGRKSIDNOLYFDPATCKQVKGQVLSIDGRNYYFDS	1525
Db	1222	VRNRVLLSDGSWYFQ	GDGVPVTVGVINGQTLFYD-ADGRQVKGQQRVIGNORYWMDK	1280
Qy	1526	GSGNMAK	1532	
Db	1281	DSGEMK	1287	

RESULT 13

B97033

uncharacterized protein, related to enterotoxins of other Clostridiales [imported] - Cl

C/Species: Clostridium acetobutylicum

C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C/Accession: B97033

R/Nolling, J.; Bretton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee

J. Bacteriol. 183, 4823-4838, 2001

A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl

A/Reference number: A96900; MUID:21359325; PMID:21359325

A/Accession: B97033

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-2817 <KUR>

A/Cross-references: UNIPROT:Q97K42; UNIPARC:UPI000000CA0A0; GB:AE001437; PIDN:AAK79053.1

A/Experimental source: Clostridium acetobutylicum ATCC824

C/Genetics:

A/Gene: CAC1079

Query Match	7.9%	Score	654.5;	DB 2;	Length	2817;			
Best Local Similarity	20.4%	Pred. No.	3.3e-21;						
Matches	438;	Conservative	271;	Mismatches	660;	Indels	775;	Gaps	114;

Qy	34	TVSAEDKVANDTTAQA	TVGVDTGQDQAT-----TNDANTNTTDDTDAQSANTNQDAQSGD	89	
Db	130	TFKNESINNEA	---PTIPKDTSKTKTSQAQTKSGSDNNIPSNNTSTNTSKNENPSNTD	186	
Qy	90	QSNQDQAKODTANTDR	NQADN-----SQTDDNQATDQATSPATDGTSVQERRAANVAT	143	
Db	187	IKTTEAPANAPIKDT	PNQSDSALAKNAUSNNLAAD---SSQTSKVYSSNNDAKPVNT	243	
Qy	144	AADQEGQTAPSEOEK	SAALSIDNVKLI	DG-----KYYYVQADGSYKKNFAITVNGOML	196
Db	244	T-----STDKAS	LNNDSSQ-----DGWVTDKGKYYV--NGVQKGFQ-SINKSIY	288	
Qy	197	YFOSD-----TG	ALSSTSTYFSQCGTTLNVDDFSHNKAYDSTAKSFELVNGVLTANS----	249	
Db	289	YFNDGSMQTCGLK	YNSYFYFASGVMLTGLQNINGTYTGFNDGDKLLTGLQAINNNY	348	

QY 1056 GD---FLPASFTGDQNAKTFYVDGTGMAYYSTSGNKAVNSFIYEGGHYFFDKDGHMVT 1112
Db 1838 NSUFYDFPIEF---NLVTTGQWINGKKYFYDFINTGAALTSYKIINGKHFEYFNNDGMVL 1893
QY 1113 GSYKAEDGNDYFLP---NGIQMDAIYQD---AQGNSYYVGRGTILYKG---DNW 1159
Db 1894 GVEKGPDPGE-YFAPANTQNNIEGQAIYQSKFELTLNGKKYFYFDNNSKAVTGWRINNE 1952
QY 1160 YPFDVFNNA-----NKTVFR-----YFDANNVMAI-GYRNM 1189
Db 1953 KYFNFENNAIAAVGLQVIDNNKYFYFNPDTAIIISKGTQVNGSRYYFDTDAIAFNGYKTI 2012
QY 1190 YGTYFYFDENGFOAKGOLLTDDKGTHYF-----EDGAMAKNFVNVGDDWYMD 1240
Db 2013 DGKHFYFSDCVVKYIGVFSTN-GFEYFAPANTYNNIEGQAIYQSKFELTLNGKKYFYD 2071
QY 1241 GNGNAVKGVPVNNQILYFNPETGVQVKGFITDAQ-----GRTSY 1281
Db 2072 NNSKAVTGWTIDSKKYFYFNTTAAATGWTIDGKKYFYFNTTAAATGWTIDGKKY 2131
QY 1282 YDANSALKSSGFTFPGSDMYAENGY---VYK---GPKQVA-ENQD----- 1322
Db 2132 FNTNT-AAIAGTYIINGKHFYFNTDGLMIGVFKGPNGFYFAPANTDANNIEGQAILY 2190
QY 1323 -----QWYFDQTTGQAKGAAKVDGRDLYFNPDSGVQV-----KGDPAFD- 1363
Db 2191 QNEFLTLNGKKYFYF-GSKAVTGWRINIINNKYFYFNPNAIAIHLCTINNDKYFSDG 2249
QY 1364 --ESG-----NTSFVHGNGDKVVGFF-----FT 1385
Db 2250 ILQNGVITERNFYFDANNESKQWTVGVFKGPNGFYFAPANTHNNNIEGQAIYQNKFL 2309
QY 1386 TGNNAWYADNNGNLVKGFEIDGKWHFDEVTGQQAALVNGQOLYFDVDSGIQVKG 1445
Db 2310 TLNGKKYFYFNDKSAVTGWTIDGKKYFYFNLTAETAAATGWTIDGKKYFYFNLTAETAAATG 2369
QY 1446 DFVTDGQ---NTSYVDVNSGDKVNG---FFTT-----GDNW----- 1478
Db 2370 WOTIDGKKYFYFNTTIASTGYTISGKHFYFNTDGLMIGVFKGPNGFYFAPANTDAN 2429
QY 1479 -----YYADGQGNLAKGKSIDNQDLYFDPATGKQVKGQVLSIDG 1518
Db 2430 NIEGQAILYQNKFLTLNGKKYFYFGSDSKAVTGWTIDGKKYFYFNTTAAVATG-WQTNG 2488
QY 1519 RNYFYDSSGNNMAKQNFVRIGQWYFGNDG 1549
Db 2489 KKYFYFNTN-STAGTYTISGKHFYFNTDG 2518
1
RESULT 15
I40884
Cytotoxin L - Clostridium sordellii
C.Species: Clostridium sordellii
C.Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
C.Accession: I40884
R:Green, G.A.; Schue, V.; Montell, H.
Gene 161, 57-61, 1995
A>Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium sordellii
A.Reference number: I40884; PMID:95369733; PMID:7642137
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-2364 <RES>
A.Cross-references: UNIPROT:Q46342; UNIPARC:UPI00000B02BE; EMBL:X82638; NID:G1000694; PI
C.Keywords: cytotoxin

Query Match 6.1%; Score 498.5; DB 2; Length 2364;
Best Local Similarity 19.3%; Pred. No. 2.6e-14;
Matches 387; Conservative 283; Mismatches 693; Indels 641; Gaps 91;
QY 34 TVSAED-----KVANDTTAQATV-----GVDTG--QDQATNDANTN----- 68
Db 437 SISNEDNMFMKITNYLKVGPAPDVRSTINLSGPGVYTGAYQDILLMFKDNSTNHLLEP 496

QY 69 -----TTDTDTADQSANT-----NQOAGSDQSNNOQAKQDNTANTDRNQ--ADNSQ 113
Db 497 ELRNFEFPKTKISQTEQETISLWSFNQARAKSQFEEYKGYFEGALGEDDNLDPQNTV 556
QY 114 TNNQATDQATSPATDGTSVQRDAAANVATAADQEGOTAP-----SQEKAASALSLD 165
Db 557 LDKDYVSKKILS-----SMKTRNKEYIHYIVOLQGDKISYEASCNLFKDPYSSILYQK 610
QY 166 NVKLIDGKYYYQADG-----SYKKNFAITVNGQ--MLYFSDTCTGALSSTST 210
Db 611 NIEGSTAIYVVADAEIKYRIPYOISNKRNIKLFIGHGKSEFNTDTFA----- 664
QY 211 YSFSGQTTNL-VDDFSSHNKAYDSTAKSFELVNGYLTAN----- 248
Db 665 -----NLDVDSLSEIETILNLAKA-DISPKYIEINLLGCNMFYSISABETYPGKL 715
QY 249 -----SWYPAGILRNGQWTEAS-----NENDLRPVL--MSWMPKQTOV----- 286
Db 716 LLKIKDRVSELMPIS-ISQDSITVSANQYEVRIINEGKREILDHSGKWINKESIIKDISS 774
QY 287 -AYVNY-----MNKYL SANET---EVTNETSOVLNKEAQSIQTKIEQKITSDNST 333
Db 775 KEYISFNPENKIIVKSKYLHSLTLLQEIARNANSSDIDLEKKVMLTECEINVASNIDR 834
QY 334 QMLRTAMEAFVAAQPKWNNSTENFNKDLQGGALLYTNSD---LTPWANSDYRLNRT 390
Db 835 QIVEGRIE-----EAK-NLTSDSIN-----YKNEFKLIESDSLYDLKHQNG 877
QY 391 TO-----ODTKKYFTGEGGEGYFLLSNVDNSNVPVQAPQLNQLHYLMWMDIVMG 443
Db 878 LODSHFISPDISKT-----ENGFRIRFINKETGNSIFIEKEIFESEYATHISKEISN 931
QY 444 DKDANFDGVRDAVDNVDNADLLQ-----VYSNFKYKNTVDSEANALAHISILEANSLND 499
Db 932 IKDTFDNVNGKLVKVNLDAHAHVNTLSAPFIOSLIEYNTTKESLNSLVAMKVQVYA 991
QY 500 NOYNEDTNGTALSIDNSSRLTSLAVLTQPGQRIDLSNLSIESVKNERANDTAYGDTIPT 559
Db 992 QLFSTGLN---TIDASKVVEL-----VSTALDETI-----DLPT 1024
QY 560 YSFVRAHSEVQTVIAKI-----VKEKIDTNSDGYTFTLDQKDAFKIYNEDMAKV 610
Db 1025 L-----SEGLPIIATIIDGVSGLAAIKELSETN-DPLLQBIEIAKIGIMAVNLTAAS 1076
QY 611 NKTYTHYNIPIAAYALLSNM-----ESVPRVYVDLYTDDGOYMAKKSPPYDAIA----- 660
Db 1077 AIVTSALGIASGFSILLVPLAGISAGISPLVNNELIQD--KATKVIDYFKHISLAETEG 1134
QY 661 --TMLQGRYAY-----VSGGQSEEVHKVNGNQNILS-----SVR 692
Db 1135 AFTLLDDKLIIMPQDDLVISEIDFNNSITLKGCEIWRAGGSGHTLTDDIDHFFSPSPT 1194
QY 693 YQGDLMASADDT-----QGTDLRSRTSLGLVTLVSNDPN-----LDLGGDSIT 732
Db 1195 YRKPWLSIVDLNIRKKEKIDFSKO---LWVLPNAPNRVFGYEMGWTPGFRSLDNDGTCLL 1251
QY 733 VNMGRAHANQAYRPIILTKGQVQSYLKD---SDTNIVKYTDAN-----GN 775
Db 1252 DRIRDHYEGQFYWRYFAFIADALITKLPRVEDTNRINLDGNTSRFIVPVITTEQIRKN 1311
QY 776 LFTTADDIKGYSTVDMSGVLA-----VWV----- 799
Db 1312 LSYSYFGGGGYSLSLSPYNNIDNLNVENDTWIDVNVKNITTESDEIOKGLIENI 1371
QY 800 --PVGAKDQ-----QDVRV----- 811
Db 1372 LSKLNIEDKNKIILNHNTHFYGDINENRFLSITFSILEDINIIEIDLVSXYKILLSG 1431
QY 812 -----AADTNQKAD-----GKSLK---TGAALDSQVIYEGFSNFPQDPANNND----ADY 852
Db 1432 NCMKLIENSDDTOOKIDHIGFNGEHQKIPYSIDNETKYNIGFI---DYSKKEGLFTAEF 1488

Db 241 VNGYLTANSWYRPAGLRLNGQTWEASNENDLRPLVLSWMPDKDTQVAYVNYNMYKLSANE 300
QY 301 TEVNTSETVDLNKEAQSQTKEIQKITSNDSTQWLRTAMEAFVAAQPKWNMSTENFNKG 360
Db 301 TEVNTSETVDLNKEAQSQTKEIQKITSNDSTQWLRTAMEAFVAAQPKWNMSTENFNKG 360
QY 361 DHLQGALLYTNISDLTPWANSYRLLNRPPTQDGTKKYFTEGEGGEGYFLLSNDVDNSN 420
Db 361 DHLQGALLYTNISDLTPWANSYRLLNRPPTQDGTKKYFTEGEGGEGYFLLSNDVDNSN 420
QY 421 PVVQAOLNQLHYLMWNGDI VMGDKDANFDGVRVDAVDNADLLQVYNYFKDNKYKVD 480
Db 421 PVVQAOLNQLHYLMWNGDI VMGDKDANFDGVRVDAVDNADLLQVYNYFKDNKYKVD 480
QY 481 SEANALAHISILEAWSLNDQYNEDTNGTALSIDSNSRLTSLAVLTQPKQGRIDLSNLIS 540
Db 481 SEANALAHISILEAWSLNDQYNEDTNGTALSIDSNSRLTSLAVLTQPKQGRIDLSNLIS 540
QY 541 ESNVKERANDTAYGDTIPTYVSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAF 600
Db 541 ESNVKERANDTAYGDTIPTYVSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAF 600
QY 601 KIYNEDMAKVNKTYTHYNI PAAYALLSNMESVPRVYVYGDLYTDDGQYNAKSPYYDATA 660
Db 601 KIYNEDMAKVNKTYTHYNI PAAYALLSNMESVPRVYVYGDLYTDDGQYNAKSPYYDATA 660
QY 661 TMLQGRIAVYSGQSEBEVHKVNGNNOILSSVRYGQDLMSADDTQGTDLRTSLGLVTLVSN 720
Db 661 TMLQGRIAVYSGQSEBEVHKVNGNNOILSSVRYGQDLMSADDTQGTDLRTSLGLVTLVSN 720
QY 721 DPNLDLGGSLTVNMGRAHANQAYRPLILGTQDGVQSYLKQSDTNI VKVYTDANGNLTFTA 780
Db 721 DPNLDLGGSLTVNMGRAHANQAYRPLILGTQDGVQSYLKQSDTNI VKVYTDANGNLTFTA 780
QY 781 DDILKGYSTVDMGYLAVVPVPGAKQGVVRAADTNQKADGKSLKTSAAALDSQVIEGFS 840
Db 781 DDILKGYSTVDMGYLAVVPVPGAKQGVVRAADTNQKADGKSLKTSAAALDSQVIEGFS 840
QY 841 NFQDPANNDADYTNKXIAENADPFKKLGITSFEMAPQYVSATDGSFLDSIIQNGYAFSDR 900
Db 841 NFQDPANNDADYTNKXIAENADPFKKLGITSFEMAPQYVSATDGSFLDSIIQNGYAFSDR 900
QY 901 YDLAMSKNNKYGSKDPLANAL KALHANGIOAIADWVPDQIYQLPGBEVTVAKRTNSYGNP 960
Db 901 YDLAMSKNNKYGSKDPLANAL KALHANGIOAIADWVPDQIYQLPGBEVTVAKRTNSYGNP 960
QY 961 TPDAYINNALYATNTKSSGSDYQAOYGGAFDLDELKAKY PDMFTVNMISTGKPIDPSTKIK 1020
Db 961 TPDAYINNALYATNTKSSGSDYQAOYGGAFDLDELKAKY PDMFTVNMISTGKPIDPSTKIK 1020
QY 1021 QWEAKYFNGTNLGKAGYVLSDDATGKFTVNEGDFLPASFTGQNAKTGPYYDGTGM 1080
Db 1021 QWEAKYFNGTNLGKAGYVLSDDATGKFTVNEGDFLPASFTGQNAKTGPYYDGTGM 1080
QY 1081 AYTSTSGNKAVNSFIYEGGHYYTFDKGHMVTGSYKAEDGNDYFYPNGIQMRDAIYQDA 1140
Db 1081 AYTSTSGNKAVNSFIYEGGHYYTFDKGHMVTGSYKAEDGNDYFYPNGIQMRDAIYQDA 1140
QY 1141 QGNSYYGRTGILYKGDNMYPFVDPNNANKTVPRYFDANNVMAIGYRNMVGYQYFDENG 1200
Db 1141 QGNSYYGRTGILYKGDNMYPFVDPNNANKTVPRYFDANNVMAIGYRNMVGYQYFDENG 1200
QY 1201 FQAKGQLLTDKETHYFDEBNGAMAKNKFVNVGDDWYMDGNGNAVKGQYVNNQILYFN 1260
Db 1201 FQAKGQLLTDKETHYFDEBNGAMAKNKFVNVGDDWYMDGNGNAVKGQYVNNQILYFN 1260
QY 1261 PETGVQVKGQFITDAAGRTSYDANS GALKSSGFFTPNGSDWYTAENGYYVYKGFQVAEN 1320
Db 1261 PETGVQVKGQFITDAAGRTSYDANS GALKSSGFFTPNGSDWYTAENGYYVYKGFQVAEN 1320
QY 1321 QDQWYYFDQTTGQAKGAACVGRDLYFNPDGSGVQVKGDFATDSEGNSTFYHGDNGDKVY 1380

Db 1321 QDQWYYFDQTTGQAKGAACVGRDLYFNPDGSGVQVKGDFATDSEGNSTFYHGDNGDKVY 1380
QY 1381 GGFFTTGNNAMYYADNNGNLVKGFOEIDGKWTHTFDEVTGQQAAGALVNGQQLYFDVDSG 1440
Db 1381 GGFFTTGNNAMYYADNNGNLVKGFOEIDGKWTHTFDEVTGQQAAGALVNGQQLYFDVDSG 1440
QY 1441 IQVKGDFVTDGOGNTSYDVNSGDKKVGFFTTGNNAMYYADGQGNLAKGRKSIDNQDLY 1500
Db 1441 IQVKGDFVTDGOGNTSYDVNSGDKKVGFFTTGNNAMYYADGQGNLAKGRKSIDNQDLY 1500
QY 1501 FDPATGKQVKGQVLSIDGRNYYFDSGSGNMAKNRFRVIGDQWIFYFGNDGAAATNL 1554
Db 1501 FDPATGKQVKGQVLSIDGRNYYFDSGSGNMAKNRFRVIGDQWIFYFGNDGAAATNL 1554
RESULT 2
Q9WXJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE GTF-S.
GN Name=gltf;
OS Streptococcus criceti.
OG Plasmid pAMI.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
RN NCBI_TaxID=1333;
RX [1]_TaxID=1333;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSSP; P06654; IMPE.
DR GO; GO:0002250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9C601FC14 CRC64;
Query Match 53.1%; Score 4375; DB 2; Length 1512;
Best Local Similarity 55.6%; Pred. No. 1.6e-182;
Matches 882; Conservative 188; Mismatches 393; Indels 122; Gaps 27;
QY 1 MEKKLHYKLHKYKHHVTTIAVASIGLVSLVGAG---TVSAEDKVANDTTAQTAVGVDTG 56
Db 1 MEKKLHYKLHKYKHHVTTIAVASIGLVSLVGAG---TVSAEDKVANDTTAQTAVGVDTG 52
QY 57 QDQATTNDANTNTTDTDTA---DQSANTNQDQAGSDQSNNOQQAQKQDTANTDRNQADNSQ 113
Db 57 QDQATTNDANTNTTDTDTA---DQSANTNQDQAGSDQSNNOQQAQKQDTANTDRNQADNSQ 113
QY 53 -EQNTSASQNEKVDSDAAQATDAKATSEQAASVSDTNS- - - - -TQVQNVQDV 102
Db 53 -EQNTSASQNEKVDSDAAQATDAKATSEQAASVSDTNS- - - - -TQVQNVQDV 102
QY 114 TDNQ-----ATDQATSPATDGTSGVRRDAANVATADQ-----EGQTAP-----SEQKS 159
Db 114 TDNQ-----ATDQATSPATDGTSGVRRDAANVATADQ-----EGQTAP-----SEQKS 159
QY 103 SANNQPEAPAAPQA--ASNNNTATSEANTNTAVSEAAPAAENRTAEKADLSQBEA 160
Db 103 SANNQPEAPAAPQA--ASNNNTATSEANTNTAVSEAAPAAENRTAEKADLSQBEA 160
QY 160 AALSIDNKLIDGKYVYQADGSYKKNFAITVNGOMLYFSDTSGALSSSTSTSFSGQTTN 219
Db 160 AALSIDNKLIDGKYVYQADGSYKKNFAITVNGOMLYFSDTSGALSSSTSTSFSGQTTN 219
QY 161 AALSIDNKLIDGKYVYQADGSYKKNFAITVNGOMLYFSDTSGALSSSTSTSFSGQTTN 220
Db 161 AALSIDNKLIDGKYVYQADGSYKKNFAITVNGOMLYFSDTSGALSSSTSTSFSGQTTN 220
QY 220 LVYDSSSHNKAYDSTAKSPPELVNGYLTANSWYRPAAGILRNGQTVWEASNENDLRPLVLSMW 279
Db 220 LVYDSSSHNKAYDSTAKSPPELVNGYLTANSWYRPAAGILRNGQTVWEASNENDLRPLVLSMW 279
QY 221 LVYDSSSHNKAYDSTAKSPPELVNGYLTANSWYRPAAGILRNGQTVWEASNENDLRPLVLSMW 280
Db 221 LVYDSSSHNKAYDSTAKSPPELVNGYLTANSWYRPAAGILRNGQTVWEASNENDLRPLVLSMW 280
QY 280 PKDQTVAYVNTVNNKYL SANETEVNTNETS QVDLNKEAQS IQKIEQKITSNDSTQWLRTA 339
Db 280 PKDQTVAYVNTVNNKYL SANETEVNTNETS QVDLNKEAQS IQKIEQKITSNDSTQWLRTA 339
QY 281 PKDQTVAYVNTVNNKYL SANETEVNTNETS QVDLNKEAQS IQKIEQKITSNDSTQWLRTA 339
Db 281 PKDQTVAYVNTVNNKYL SANETEVNTNETS QVDLNKEAQS IQKIEQKITSNDSTQWLRTA 339
QY 340 MEAFVAAQPKWNMSTENFNKGDHQLQGGALLYTNISDLTPWANSYRLLNRPPTQDGTKKY 399
Db 340 MEAFVAAQPKWNMSTENFNKGDHQLQGGALLYTNISDLTPWANSYRLLNRPPTQDGTKKY 399

Db 340 MAAFVATQGRWNSDSQFDKNDHLOGGALLYTNKLTWEADSKYRLNLTPTFQDGKTHY 399
QY 400 FTEGEGGVEFLLSNDVNSPVVQAEOLNOLHYLMNWGDI VMGDKDANFVGVRVDAVDN 459
Db 400 SKADEYGGVEFLLANDVNSPVVQAEMLNQIHYLMNWSI VMGDKDANFVGIRVDAVDN 459
QY 460 VNADLLQVTSNPKONYKVTDSANALAHISILEAWSLNDNQYNEDTNGTALSIDNSSRL 519
Db 460 VDADTJQLYNTYFNAYGVGDKSEQAALAHISILEAWSYNDNYNQDTNGAALAMDNGRL 519
QY 520 TSLAVLTQPGORI-DLSNLI-SESUNKERANDTAYGDTIPTYSFVRAHDSEVQTVIAKI 577
Db 520 SLLYTLTRPLSERTPGSLTLIKSEYGLTDRDKYGDTPQSPYFVRAHDSEVQTVIAQI 579
QY 578 VREKIDTNSDGYFTFLDQLDKAFKINEDMAKVNKYTHYNIPAAVALLSNESVPRVY 637
Db 580 IREKIDPTDGTFTFLDQLKQAFDIYNKDNVSEKHYTHYNIPAAVALLSNESVTRVY 639
QY 638 YGDLYTDDGOYMAKSPYYDAIATMLQGRIAVYSGQSEVHK----- 680
Db 640 YGDLFTDDGOYMETKSPYYDAINTLLRARIRYAAGGQTMW-HKAYTPSAAMKAKNPDSGS 698
QY 681 VGNNOILSSVRYGQDLMSADDTQGTDLSTRTSLVTLVNSDNDPLD--GGDSLTVMNGRAH 739
Db 699 VLGNSEVLVSVRFGQDWSADDMTGGQLAKTSGMFSLIANNPELELDANEELKVNVGKIH 758
QY 740 ANQAYRPLLTGKQGVQSVLKQSDTNI VKYTDANGNLTFADDIKGYSTVDNSGYLAVWV 799
Db 759 ACQAYRPLLTGKGLQKYLNSDNTLTVAQDGFITFKGSEIKGYKQVGVNGVLSVWV 818
QY 800 PYGAKGQDVRVAADNTQKADG-KSLKTSAAALDSQVIYEGFSNFQDFANNADYTNKKIA 858
Db 819 PYGAKSDQDIRVAASATKANAKGDKSVTASQALDSQLIYEGFSNFQDFVQKDAQYTNKKIA 878
QY 859 ENADFPKGLTSPFMAPOYVSATDGSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDIA 918
Db 879 ENTDLFKAWGVTSFMAPOYVSATDGTFLDSIIQNGYAFSDRYDLAMSKNNKYGSKEDIA 938
QY 919 NALKALHANGIOAIAADWPQIYQLPGREVVTAKTNSYGNPTFPAYINNALYANTKSS 978
Db 939 NALKALHAGIOAIAADWPQIYQLPGREVVTAASRVNDYGRVKIDQPMVKLYLANTKSS 998
QY 979 GSDYQAOYGGAFLELAKARYPDMFTVMNIMSTGKPIDPSTKIKQWEAKYFNGTNVLGKAG 1038
Db 999 GKDFQAKYGGFEFLAELOKQYPEMFTAKMISTGKPIDSSVKLKEWSAQYFNGTNVLGRGTD 1058
QY 1039 YVLSDDATGKYFTVNEGDFLPASTGDNQAKTGFPYDGTGMAYYSTSGNKAVNSFIYEG 1098
Db 1059 YVLSDEGTGKYFTVNEKGFEFLPAVLTDGKEAKTGFPYDGTGMAYYSTSGNKAVNSFIYEG 1118
QY 1099 GHYYFDDKGMVTSYKAEDNDYFLPNGIQMRDAIYQDAQNSYVYGRGILYKG-- 1156
Db 1119 GNTYFDTGHWVTP-NGINTKFFYFLPNGVMLKDAVMEDDRGSVYVYGTGMVYKXSR 1177
QY 1157 -DNWYFFVDPNNANKTVRYFDANNVMAIGYRNMVGTQYVFPENGFOAKGOLLTDKQ-T 1214
Db 1178 NNEWFAMTD--SKGQLRFRHFDNYGPMVGLVTIHNQYVYDEEGFQVKGDFVTDKAGQT 1235
QY 1215 HYFEDNGAMAKNFVNVGDDHYNDMGNGNAVKGOYPVNNQILYFNPNPETGVQVQGFITD 1274
Db 1236 RYFDKNTGNLVKGQFENQNGHWYSDQGLI AKGAQTIKGOKLYFDAKTGAQVKGDFVTD 1295
QY 1275 AQGRTSYYDANSALKSSGFFTPNGSDWYIA-ENGYYVYKGFQVNAENQDQWYFQDTGK 1333
Db 1296 KQNTFFYSGDGTDLAVSTFFSTGNNAWFIADENGHV----- 1332
QY 1334 QAKGAAYDGRDLYFNPNDSGVQVKGDFATDESGNTSFYHGDNGDKVVGGEFTT--GNNAW 1391
Db 1333 -AKGEKTINGKLYFDTKGQAKGRFVRDAGK-IRFTDADTGALVTVNSFLETKAGSNQW 1390
QY 1392 YYADNNGNLVKGFBQIDGKWHFDEVTCQAKGAALVNGQOYLFDVDSGIVQVKGDFVTDG 1451

Db 1391 YYMGADGYAVRGHQTITQSRHMYFDAETQQAQKIV-----VTDA 1429
QY 1452 QGNTSYDYNVDGKKNVGFFTTGDNAWYADGOGNLAKEGRKSIDNODLYFDPATGKQVKG 1511
Db 1430 NERKIFYDANTGDRVNVQFVLV-NGSWYFFGYDGAATGFRDIRGOHLYFNP-DGTQAKG 1487
QY 1512 QLVSIDGRNYYFDGSGGNMAKRFV 1536
Db 1488 TTVKIDNRIYTFDADSGELTSVRYI 1512
RESULT 3
Q56CX8 9STRE PRELIMINARY; PRT; 1506 AA.
AC Q56CX8;
DT 10-MAY-2005 (TREMBLrel. 30, Created)
DT 10-MAY-2005 (TREMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TREMBLrel. 30, Last annotation update)
DE Glucosyltransferase-T.
GN Name=gtf;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
RN NCBI_TaxID=1310;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B13N;
RA Kuwahara N.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY966490; AAX76986.1; -; Genomic_DNA.
KW Transferase.
SQ SEQUENCE 1506 AA; 167016 MW; 94B882EF2C17C451 CRC64;
Query Match 52.9%; Score 4360.5; DB 2; Length 1506;
Best Local Similarity 54.7%; Pred. No. 6.7e-182;
Matches 869; Conservative 202; Mismatches 349; Indels 169; Gaps 25;
QY 1 MEKLLHYKLHKVKKHWITIAVASIGLVSLVGAG---TVSADKQVANDTTA-QATYGVDT 55
Db 1 MERKLLHYKLHKVKKHWITIAVASAGLIVGAGSLQTVSADDLAKEQAASQQKAAANQ 60
QY 56 QGDQATTNDANT---NTTDTDTADQSANQDQAGSDQSNQDQAK-QDTANTDRNQADN 111
Db 61 NEDEVASDAADTASAKATSEKVVQSDTNS--TNQVETKQASAKESADAVAKAPQ 117
QY 112 SQTNNQATDQATSPATDGTSVQRDAANVATA--ADQEGQTAPSEQEKSAALSIDNVKL 169
Db 118 A-----GPAIT-SQVASESSVAPSKADKAAAGSVSQNEEALSLANIKK 164
QY 170 IDGKYVYVOADGSYKKNFAITVNGOMLYFSDTGLSSTSTVSQGTNLVDDFSHNK 229
Db 155 IDGKYVYVADGSYKKNFAITVDCOMLYFDARTGALSSTSTVSQGLTPIVSDFSVNNK 224
QY 230 AYDSTAKSFELVNGYLTANSWYRPAIGILRNQGTWEASNENDLRPVLMNSWPKDQTVAYV 289
Db 225 AFDSEKSFELVDGYLTAESWYRPAKILLENKGTWVDSKETDLRPVLMNSWPKDQTVAYL 284
QY 290 NYMKYLSANETEVNTNETSQVDLNKBAOSIQTKIEQKITSNSTOWLTAMEAFVAAQPK 349
Db 285 NYMSKALGGKE-EFTTETSQTLNTAAELIOAKIEARVSKEGQTKWLEMAAFVATQSR 343
QY 350 WNMSTENFNKGDHLOGGALLYTNSDLTTPWANSYVLLNRTPTQODGTKKYFTEGEGGYE 409
Db 344 WNKQSEYQKADHLOGGALLYTNNNLTWANSWRLNRTPTROQKTHYSKADYKGYE 403
QY 410 FLLSNDVNSNPVQAEOLNOLHYLMNWGDI VMGDKDANFVGVRVDAVDNVDNADLLQVYS 469
Db 404 FLLANDVNSNPVQAEMLNQIHYLMNWSI VMGDKDANFVGIRVDAVDNVDNADLLQVYT 463
QY 470 NYFKDNYKVTDSANALAHISILEAWSLNDNQYNEDTNGTALSIDNSSRLTSLAVLTQKP 529
Db 464 NYFNSVYGVNKSEQAALAHISVLEAWSYNDNYNQDTNGAALAMDNGRLSLLYTLTRPI 523

```
QY 530 GORI-DLSNLI-SESVNKERANDTAYGDTIPTYSFVRHDSVOTVIKIVKEKIDTNSD 587
Db 524 NERTPGMSTLIKSEYGLTDRKNDKYGDTPQSPYFVRHDSVQTVIAQIIKEKIDPTD 583
QY 588 GYTFILDQLDKAFKINEDMAKNTKTYHNTYPAAYALLSNMESVPRVYVYGLYDITDQ 647
Db 584 GYTFILDQLKQAFKINEDMAKNTKTYHNTYPAAYAVMLSNMESVTRVYVYGLDITDQ 643
QY 648 YNAKSPYYDAIATMLQGRYAVSGQSEE-----VHKVNGNQNILSV 691
Db 644 YNAKSPYYDAINTLLRARIRYAAGGOIMEHNSYKPSAAMKAHPDAGNVLGNSVLSV 703
QY 692 RYGODLMSADDTQGTDLSTSLVTLVSNPDLIG-GDSLTVNMGRAHQAAYRPLILG 750
Db 704 RFQGDVMSADDTGGLAKTSGMFTLISNPELELDVNEEIKVNVGKIHAGQAYRPLLT 763
QY 751 TKDQVQSYLKSDSTNIVKVTYDANGNLTFADDTIKGYSTVDMGSLAVVWVPGAKQDVR 810
Db 764 TKGLQKYLNDSDTKLTADKDGFTTFKSGEIKGYQVEVNGYLSVWVPGAKADQIR 823
QY 811 VAADTNQKAD-GKSLKTSAAALDSQVIYEGFSNFPQDPANNADYTNKKIAENADFFKGLGI 869
Db 824 VAPSTAAGKEKARTYASQALESQLIYEGFSNFPQDPVQKDSQYTNKKIAENTDLFKAMGV 883
QY 870 TSFEMAPQVVSATDGSFLDSIIQNGYAFPSDRYDLAMSKNKKYGSKODLANALKAHANGI 929
Db 884 TSFEMAPQVVSATDGSFLDSIIENGAYFTDRYDLAMSKNKKYGSKEDLANALKALHAAGI 943
QY 930 QAIADWVPOIYQLPQEEVVTAKRTNSYGNPTFDAYINNALYATNTKSSGSDYQAYGGA 989
Db 944 QAIADWVPOIYQLPQKEVVTASRVNDYGRVVDQPLVEKLYLANTKSSGKDPQAYGGE 1003
QY 990 FLDELKAKYPMFTNMISTGRPIDPSTKIKOWEAKYFNGTIVLKGAGYVLSDDATGY 1049
Db 1004 FLAELQKKYPEMFTTMTSTGKTIDPSVKLKEWSAKYFNGTIVLDRGTDYILSDEGTGY 1063
QY 1050 FTVNENGDFLPASFTGDQNAKGFYDGTGMAYYSTGNKAVNSFTYEGGHYVYDQKH 1109
Db 1064 FTVNEKGDFLPASLTGNKDAKGFYNDGKGIYVYTTAGNKARSAFYTEAGNTYFYDYTH 1123
QY 1110 MYTGSYKABDGNDDYFLPNGIOWRDAIYQDAQNSVYVYGRGTGLYKG--DN-WYFVDPN 1166
Db 1124 MYTGP-NVINTKPYFLPNGIMLKDAIKQDEKRSVYVYKTVGMYKGRDNEFWAMTDSK 1182
QY 1167 NANKTVRYFDANNVMAIGYRNMYGQTYFYFDENGFOAKGOLLTDKG-THYFDENGAWA 1225
Db 1183 GQWR--FRHFDRYGFMSIGLVTINQNVYDENGFOVKGFEVTDODGQTRFYDQSGNLV 1240
QY 1226 KMKFVNVDGDDWYMDGNGNAVKQYFPVNNQILYFNPETGVQVKGQFIDAQGRTSYYDAN 1285
Db 1241 KGQFLNKDGNWYLD----- 1255
QY 1286 SGALKSSGFFTPNGSDWYVAENGYYVYKGFQKVAENQDQWYFDTTGTQKQAKGAVDGRD 1345
Db 1256 -----DQ--GLVAKGQATIKGK 1271
QY 1346 LYFNPDSGVQVKGDFATDSGNTSFYHGDNGKWKVGGFTTGNNAWYADNNGNLVKGPQ 1405
Db 1272 LYFDTKTGVQVKGDFVTDKDGNTFFYSGDTGDLILGQFFSTGNNAWYADENGHV----- 1326
QY 1406 EIDGKWHYFDEVTGQAKAALVNGQOLYFDVDSGLQVKGDFVTDQGNSTYVDVNSGDK 1465
Db 1327 -----AKGAKTIRGQKLYFDTKTGOQAKGRFIRDDKG-VRYYDADTGL 1369
QY 1466 KVNQFFTT--GDNWYVYADGQNLAKGRKSIDNQDLYFDPATGKQVKGQVLS-IDGRNYY 1522
Db 1370 VTNAFLETAGSNQWYMGADGVAVKGNQTIKNQHYFPAETGQQAAGLIIVTDANGKYP 1429
QY 1523 FDSGSGNMAKRPVRIGDQWYFNGDAA 1551
Db 1430 YDTFTGSRVNVQFVLVNGWNYFPGYDAA 1458
```

```
RESULT 4
Q00599_STRSL
ID Q00599_STRSL PRELIMINARY; PRT; 1599 AA.
AC Q00599;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-FEB-2005 (TREMELrel. 29, Last annotation update)
DE Glucosyltransferase S.
GN Name=gtfK;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=93381463; PubMed=8371114;
RA Giffard P.M., Allen D.M., Milward C.P., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and evolution of the gtf genes of oral streptococci.";
RL J. Gen. Microbiol. 139:1511-1522(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least tow glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
RL J. Gen. Microbiol. 137:2577-2593(1991).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX Jacques N.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX Giffard P.M., O'Connor D.M., Milward C.P., Simpson C.L., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and the evolution of the gtf genes of oral streptococci.";
RL Mol. Microbiol. 0:0-0(1992).
DR EMBL; Z11872; CAA77898.1; -; Genomic DNA.
DR EMBL; Z11873; CAA77901.1; -; Genomic DNA.
DR FIR; S22737; S22737.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1599 AA; 176480 MW; 24B77869E152B707 CRC64;

Query Match 48.7%; Score 4014; DB 2; Length 1599;
Best Local Similarity 51.7%; Pred. No. 9.8e-167;
Matches 851; Conservative 215; Mismatches 430; Indels 150; Gaps 40;

QY 1 MEKLYHKLHKVKKHWVTIAVASIGLSLVGAGTVSA-----EDKVANDTTAQATVGD 54
Db 1 MENKRYKLHKVKKQWVTLAVASVALATIVGGSVATSSLASAEETNNSGSPSTTTVGEN 60
QY 55 TG-----QQATTNDANT-NTTDDP-----TAPQSANTN-QDAQSGDSNQQAQKQDTAN 103
Db 61 TNPVVEKEVGTTEVANTSNTATTTERAEVTADKPAGTTVQPNSGT-----T 106
QY 104 TDRNOA--DNSQTDNNQATDQATSPATDGTSPQREDAANVATAA--DQEQQTAPSEOEKA 160
Db 107 SDRAAAVEVEAKPETTAKPEVATKDETATTS---EVAANAGVAAPTTEKSKELSEAIEKA 163
QY 161 ALSLDNVKLI-IDGKYVYVQADGSYKKNFAITVNGQMLFYDSDTGALSSSTYSFSQGTTN 219
Db 164 AVSLDNIKKDKGYLLLEDGSHKKNFAITVNGQVLYFD-ENGALSSSTYSFTQETTN 222
```

```
QY 220 LVDDFSSHNKAYDSTAKSPELVNGYLTANSWYBPAGILNRGQWTEASNENDLRPVLMSW 279
DB 223 LVTDFTKNAAYDSTKASPELVGDLTADSWYRPKBIELEAGTTWKASTEKDFRPLMSW 282
QY 280 PKDQTOVAVYNNYKYL--ANETE--VTNETSQVDLNKEAQSOTKIEQKISDSNSTOWL 336
DB 283 PKDQTOVAVYNNYKYL--ANETE--VTNETSQVDLNKEAQSOTKIEQKISDSNSTOWL 342
QY 337 RTAMEAFVAAQPKWNMTNFNGDHLQGGALLYTNSDLTPWANSYRLLNRPRTQDGT 396
DB 343 RQSEAFVQDQKWNINSBPGK-EHFQKALLFVNSDSTKANSYRKLNTQATSYIKN 401
QY 397 KYFTFGSGGVEFLSLNDVNSNPVQAEQLNLHLYLMNWGDIVNG--DKDANPDGVRV 454
DB 402 HK-IVNGSDGGVEFLSLNDVNSNPVQAEMLNQLYFNMWGOIVFGDKDKDAHFPGIRV 460
QY 455 DAVDNVNADLQVSNYKDYKVTDSANALAHISILEAWSLNDQYNEDTINGTALSID 514
DB 461 DAVDNVSDMLQLVSSYMKAAKYKVNSEARALANISILEAWSHNDPPYYNEHNTAALSMD 520
QY 515 NSRLTSLAVLT-----KOPGOR-IDLSNLISESV--NKERANDTAYGDT-IPTYSFVRA 565
DB 521 NGLRLSVHGLRTPVTKGTGARNASMKDLINGGYFGLSNRAEVTSYDQLGFATYLFVRA 580
QY 566 HDSEVQTVIAKIVEKIDTNSDGYTFTLDQKDAFKIYNEDMAKNKTYTHYNIIPAAYAL 625
DB 581 HDSEVQTVIADIISKIDPTDGTFTLDQKAPDIYNADMLKVDKVTHTSNIPAAYAL 640
QY 626 ILSNMSVPRVYGDLYTDGQYMAKSPYYDAIATMLQRIATYVSGGSEBHVKNV-- 683
DB 641 MLQTGAATRVYGDLYTDGQYMAKSPYFDITLLKARPKYVAGQTSYTHNLADG 700
QY 684 -----NNQILSVRYGDLMSADDTQGTDLRTSLGLVTLVSDNPLDLG-GDSLTVNMG 736
DB 701 VSSAKDNKEVLVRYGDLMSKTDTEGGKYGENSGMLTIANPNPKLADGETTFVNMG 760
QY 737 RAHANQAYRPLILGTGQVQSYLKDSDTNIVKYTDDANGNLTFADDIKGYSTVDMSGYLA 796
DB 761 AAHKQAYRPLILGTGQVQSYLKDSDTNIVKYTDDANGNLTFADDIKGYSTVDMSGYLS 820
QY 797 VWPVGAQDQVRYVAADTNQADG-KSLKTSAAALDSQVIYEGFSNFQDPANNADYTNK 855
DB 821 VWPVGAATDQNLAKPSTKAYKEGDKVYSSAALEAQVIYEGFSNFQDPVFKEDSYTNK 880
QY 856 KIAENADPPKLGITSEFENAPQVVSATDGSFELDSIIQNGVAFPSDRYDLAWSKNKYGSKD 915
DB 881 LIAANADLPKSGWITSEFELAPQVVSATDGSFELDSIIQNGVAFPSDRYDLAWSKNKYGSK 940
QY 916 DLANALKALHANGIOAIDWVPDQIYQLPGEVEVTAKTNSYCNPTFDAYINNALLYATNT 975
DB 941 DLRLALKALHKOGIOVIADWVPDQIYQLPGEVVTATRTDTHGKVLDDTSLVNLKLYVNT 1000
QY 976 KSSGSDYQAYGGAFLDELKAKYPMFTVNMISTGKPIDPSTKIKQWEAKYFNGTNVLGK 1035
DB 1001 KSSGSDYQAYGGAFLDELKAKYPMFTVNMISTGKPIDPSTKIKQWEAKYFNGTNVLGK 1060
QY 1036 GAGYVLSDDATCK-YFTVNENGDPLPASFTGDNQAKTGYFDGTGMAYYSTSGNAKVNPF 1094
DB 1061 GSDYVLSDDATCK-YFTVNENGDPLPASFTGDNQAKTGYFDGTGMAYYSTSGNAKVNPF 1117
QY 1095 IYEGGHYVYFDKDHMTVSKAEADGNDYFPLNGIQMRDAIYQDAQGNSSYYVGRGTILY 1154
DB 1118 VTYNGKQYVYFNDKGLYVTE-QTIDGSNYFFLPNGVMFTDGYRKNAGKQSLYVGSGLIT 1176
QY 1155 KGDNNYPPV--DPNNANKTVRYFDANNVMAIYGRNMVQTYFYFDENGQAKQOLL----- 1208
DB 1177 TOTGHWKVTXKDDSGKEEFYQYFFKGGIMATGLTEVEGKEYFYDNGYQAKGVFVPTKD 1236
QY 1209 -----TDDKG-----THYFEDDNGAKAKNPF 1229
DB 1237 GHLMPFGDGSBERKYSGFPEQDGNWYANDKGYVATGFTKVGKQNLVFNK-RGVQVQNR 1295
```

```
RESULT 5
Q00600 STRSL PRELIMINARY; PRT; 1518 AA.
AC Q00600;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-FEB-2005 (TRENBLrel. 29, Last annotation update)
DE Glucosyltransferase-I.
GN Name=gtfJ;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least tow
glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
RL J. Gen. Microbiol. 137:2577-2593(1991).
RN (2)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z11873; CAA77900.1; -; Genomic_DNA.
DR EMBL; M64111; AAA26896.1; -; Genomic_DNA.
DR PIR; A44811; A44811.
DR HSSP; P06653; 1GVN.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1518 AA; 167730 MW; DAA41F717098B59A CRC64;
```

```
Query Match 46.4%; Score 3818; DB 2; Length 1518;
Best Local Similarity 50.1%; Pred. No. 3.3e-158;
Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
QY 1 MEKLLHYKLHKVKKHVTIIVASIGLSLVG-----ACTVSA---EDKVANDTTAQATVG 52
DB 1 MENKHYKLHKVKKHVTIIVASVALATVIGLSVTTSSVSADETQDKTVTQNSGTTAS 60
QY 53 VDTGQDQATTNDANTNTTDTDTDQASANTNQDQAGSDQSNQDQAKQDTANTDRNQADNS 112
```

```
Db      61  LVTSPEATKADRTNTKADVLTPAKETNAVETAT-TTNTQATAAATATTADVA-VA 118
QY      113  QTDNNQATQATSPATDGTSGVQRDA---ANVATAADQEQTAPSPQEKSAALSLSNVKL 169
Db      119  AVPNKEAVVTDAPAVTTEKABEQPATVKAENVNTEVKAPAALAKDSEVEAALSLSNIKN 178
QY      170  IDGKYIYVQADSGYKKNFAITVNGQMLYFDSDTGALSGTSTYSFSOGTTNLDVDFSSHNK 229
Db      179  IDGKYIYVNEGDSGHKENFAITVNGQLLYFKGD-GALTSSTSTYSFTPGTTNIVDGFSINNR 237
QY      230  AYDSTAKSRELNVNGYLITANSVRPAGILRNGOTWEASNENDLRPVLMSWMPDKDQTOVAV 289
Db      238  AYDSSASRELIDGYLTADSVTPASIIKDGVTWQASTAEDFRPLLMWMPNVDITQVNYL 297
QY      290  NYMNYKLSANETVNTSETSOVLNKEAQSIQTKIEQKITSDNSTOWLRTAMEAFVAAQPK 349
Db      298  NYMSKVFNL-DAKYSSTDQOETLKVAAKDIOIKIEQIOAKESTQWLRETI SAFVKTPQ 356
QY      350  WNMSTENFKG---DHLGGALLYTNSDLTPWANSYRLNLRTPPODGT--KKYPTGG 404
Db      357  WNKETENYSGGGEDHLOGGALLVYNDSTRTPWANSYRLNLRTPATNQTGTIDKSIILDEQS 416
QY      405  E---CGYFELLSNDVNSNPVQAEOQLNOLHVLNMGDIUMGDKDANFDGVRVDAVDNV 460
Db      417  DPNHMGDFLLANDVDLSNPVVQAEOQLNQIHYLMNMGSIUMGDKDANFDGIRVDAVDNV 476
QY      461  NADLLQVSNYKNDYKVTDSANALAHISILEAWSLNDNQYNEEDTNGTALSIDNSSRLT 520
Db      477  DADMLQLYTYNPREYGVNKSANALAHISVLEAWSLNDHNYNDKTDGAALAMENKQRLA 536
QY      521  SLAVLTK-----QPGQRIDLSNLISESVNKERAND-----T 551
Db      537  LLFSLAKPIKERTPAVSPLYNNTFNTTQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNE 596
QY      552  AYGDTPPTYSFVRAHSEVQVTIAKIVKEKIDTNSDGYTFLLDQDKAFKINEDMAKYN 611
Db      597  KYGDASGNVYFIRAHNNQVDIIAEIIEKINPKSDGFTITDAEMKQAFIYNKMLSSD 656
QY      612  KTYTHYNIPAAVALLSNSESVPVYVGYDLYTDGQYMAKSPYDAIATMLQRIAYVS 671
Db      657  KKYTLNNIPAAVAVLQNMETITRVYVGYDLYTDGCHYMETKSPYDITVNLKSRIKYVS 716
QY      672  GGQSEVH-----KVGNNQILSSVRYGQDLMSADDTGTDLSRTSGSLVTLV 718
Db      717  GGQAQRSYMLPTDGMKDNDSVELYRTNEVYTSVRYGKDINTANDTEGSKYSRTSGQVTLV 776
QY      719  SNDPNLDLGDGS-LTVNMGRAHANOAVRPLILGTDKGVSQYLKQSD--TNIVKYITDNG 774
Db      777  ANNPKNLNDQSAKLANVEMGKIHANQYRALIVGTADGINKNFTSDADAIAGVYVYKETDSNG 836
QY      775  NLTFTTDDIKGYSTVDMISGLAVVPVPGAKGDQVRVAADTNTQKADGK-SLKTSAALDSQ 833
Db      837  VLTFFGANDIKGYETFDMSGFVAWVPVPGASDNQD IRVAPSTEAKKEGELTLKATEAYDQ 896
QY      834  VYIEGFSNFQDF--ANNRDADYTNKKIAENADFPKKLGITSEFEMAPYVSATGSLFDSII 891
Db      897  LIYIEGFSNFQTI PDGSDPSVYTNKRKAENVDLPFKSMGVTSEFEMAPQFVSADDTGLFDSVI 956
QY      892  QNGYAESDRDYDLMSKNNKYGSKDDLANALKALHANGIOAIADWVPDQIYOLPGEVWTA 951
Db      957  QNGYAFADRYDLMSKNNKYGSKEDLDALALKHAGIOAIADWVPDQIYOLPGEVWTA 1016
QY      952  KRTNSYGNPTFDAYINNALYATNTKSSGSDYOQYAGAFDELKAKYPMFTVMNLSGK 1011
Db      1017  TRTDGAGRKIADAIIDHSLYVANSKSGKDYQAKYGEFEFLAELKAKYPENFKVNMISGK 1076
QY      1012  PIDPSTYKIKOWEAKYPNGTNVLGKAGYVLSDDATGKYFTVNEGSGDFLPASFTGDQNAKT 1071
Db      1077  PIDDSVKLQKWAKEYENGTNVLERGVGVLSDSATGKYFTVTKEGNFIPQLTGTGKEKIT 1136
QY      1072  GFYYDGTGMAYYSTGKNKAVNSFIYEGGHYVYFDKDGHMVVTGYSKAEADGNDYY-FLPNGI 1130
```

```
Db      1137  GFSSDGKGIITYPGTSGTQAKSAFVTFNGNTYYTFDARGHMVNTSEYSPNGKDVYRFLPNGI 1196
QY      1131  QMRDAIYODAQNYYGYGRTGILYKGDNMWYPFVDPNNANK-----TVFRYFDANVMA 1183
Db      1197  MLSNAFYIDANGNTLYLXNSKGQYKGG--YTKFDVSETDKDGKESKVVKFRYFTNEGUMA 1254
QY      1184  IGRNMYGOTYYFDENGFOAKQQLTDDKGTHTYFEDINGAMAKNFVNVDGDDYYWDGNG 1243
Db      1255  KGVTVIDGTQYFGBDGFQAKDKLVTFKGTYTFYFAHTGNGIKDTWRNNGKWIYFEDANG 1314
QY      1244  NAVKGOYPVNNQIILYFNPETGVQVKGQFITDAQGRTSYDANDSANGALKSSGFFTPNGSDWY 1303
Db      1315  VAATAQVINGQKLFPN-EDGSQVKGGVVKNADGTYSKYKEGFGELVTNEFTTIDGNNVY 1373
QY      1304  YAENGYYVYKGFQKQVAENQDQWYFYDQTTGKQAKGAKVDRDLYFNPDPGQVKGDFATD 1363
Db      1374  YA-----GANGKTVTGAQVINGQHLFYFAD--GSQVKGGVWKN 1409
QY      1364  ESGNTSFYHGDNDKVKVGVGFFTTGNNAWYADNNGNLVKGFOEIDGKWTHFDEVTGQOAK 1423
Db      1410  ADGTYSKYNASTGERLTNEFTTIDGNNWYIYGANGKSVTGEVKI----- 1453
QY      1424  GAALVNGQOOLYFVDSGQVKGDFVTDGQGNTSYDVSNGDKKXNGFFTTGDNAWYYADG 1483
Db      1454  -----GDDTYFFAKDGQVKGQTVSAGNGRISYYIYDGSGRKRAVSTWIEIQGVYVYFDK 1507
QY      1484  QG 1485
Db      1508  NG 1509

RESULT 6
Q54178 STRGN PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfG;
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RX MEDLINE=96157084; PubMed=8586195;
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
phase variants.";
RL Dev. Biol. Stand. 85:309-314(1995).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Challis;
RA Minick P., Vickerman M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U12643; AAC43483.1; -, Genomic_DNA.
DR PIR; B41898; B41898.
DR HSSP; P06653; 1HCX.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_I; 5_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 45.1%; Score 3711.5; DB 2; Length 1577;
Best Local Similarity 48.0%; Pred. No. 1.6e-153;
Matches 788; Conservative 226; Mismatches 461; Indels 167; Gaps 38;
```


"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans.", *Let. 161:331-336* (1998).

[3]

NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

STRAIN=UA159 / ATCC 700610 / Serotype c;

MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;

Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P., Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A., Ferretti J.J.;

"Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.", *Proc. Natl. Acad. Sci. U.S.A.* 99:14434-14439 (2002).

-!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.

-!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-fructose + (1,6-alpha-D-glucosyl) (n+1).

-!- SUBCELLULAR LOCATION: Secreted.

-!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

-!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.

-!- SIMILARITY: Contains 6 cell wall binding repeats.

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

EMBL; M29236; AAA26895.1; -; Genomic DNA.

EMBL; D88653; BAA26103.1; -; Genomic DNA.

EMBL; D88656; BAA26107.1; -; Genomic DNA.

EMBL; D88659; BAA26111.1; -; Genomic DNA.

EMBL; D88662; BAA26115.1; -; Genomic DNA.

EMBL; D89979; BAA26121.1; -; Genomic DNA.

EMBL; AB014932; AAN58619.1; -; Genomic DNA.

HSP; P06653; IGVM.

InterPro; IPR002479; Cell wall bd put.

InterPro; IPR003318; Glyco_hydro_70.

Pfam; PF01473; CW_binding_1; 4.

DR Pfam; P02324; Glyco_hydro_70; 1.

DR Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal; Transferrase.

FT SIGNAL 1 ? Potential.

FT CHAIN ? 1462 Glucosyltransferase-S.

FT REPEAT 1232 1295 1.

FT REPEAT 1296 1359 2.

FT REPEAT 1360 1423 3.

FT REGION 1232 1423 3 X 63 AA approximate tandem repeats.

FT VARIANT 10 10 Y -> H (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT VARIANT 19 19 I -> V (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT VARIANT 58 58 K -> E (in strain MT4467).

FT VARIANT 68 68 A -> S (in strain MT4239 and strain MT4245).

FT VARIANT 81 81 A -> T (in strain MT4251 and strain MT8148).

FT VARIANT 113 113 T -> I (in strain MT4239 and strain MT4245).

FT VARIANT 122 122 A -> V (in strain MT4239, strain MT4245 and strain MT8148).

FT VARIANT 132 132 A -> S (in strain GS-5 and strain MT4467).

FT VARIANT 135 135 A -> V (in strain MT4245).

FT VARIANT 137 137 A -> T (in strain GS-5, strain MT4239,

strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

V -> L (in strain MT4239).

E -> N (in strain MT8148).

E -> D (in strain MT4239, strain MT4245 and strain MT4251).

D -> N (in strain MT4239, strain MT4245 and strain MT4251).

Q -> H (in strain MT4245).

D -> N (in strain MT4239 and strain MT4251).

E -> K (in strain MT4239).

V -> F (in strain MT4239).

F -> L (in strain MT4239, strain MT4251 and strain MT4467).

KKKYTO -> EKEYTL (in strain MT4251).

A -> S (in strain MT4239).

TDQSEA -> ADKGNDS (in strain MT4251).

TDQGS -> ADKGN (in strain MT4239 and strain MT4245).

T -> A (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

D -> Y (in strain MT4251).

E -> K (in strain MT4245 and strain MT4251).

LG -> IR (in strain MT4251).

G -> R (in strain MT4245).

G -> R (in strain MT4239).

H -> Q (in strain GS-5).

S -> N (in strain MT4239).

Y -> C (in strain MT4251 and strain MT4467).

F -> L (in strain MT4467).

Q -> P (in strain MT4245).

K -> T (in strain MT4245).

N -> D (in strain MT4245).

G -> D (in strain GS-5 and strain MT4467).

G -> R (in strain GS-5).

R -> K (in strain MT4467).

RYDKNSGMVKNVTVLANGRRIGDRNGIARY -> VY R (in Ref. 1).

SQ SEQUENCE 1462 AA; 163388 MW; CE4A379C4D708645 CRC64;

Query Match 44.4%; Score 3660; DB 1; Length 1462;

Best Local Similarity 49.5%; Pred. No. 2.5e-151;

Matches 760; Conservative 219; Mismatches 429; Indels 126; Gaps 33;

QY 1 MEKLLHYKLHKVKKHWTIAVASIGLVSLVGAGTVSAEDKAVNDTTAQTAVGYDTGQ--- 57

DB 1 METKRYKMYKVKHWTIAVAS-GLITL---GTTLSSVSAETQQTSDKVVTKQKSD 56

QY 58 DQATNDANTNTDTTADQSANNTQDQAGSQSNQDQAKQDT-----ANTDR 106

DB 57 DKAASESQTDAPKT---KQATQETQQAQS-QANVADTSTSTKTPSQNTTQANSDD 111

QY 107 NQADNSQTDNQATQATSPATDGTGVQRDANVATA-ADQEGQTAPSEQKSAAL-SL 164

DB 112 KTVTNTKSEEAQTSERTKQAEQAQATASSQALQAKAELTKQRTAAQENKPNVDLAAI 171

QY 165 DNVKLIDGKYYVQADSGYKKNFALTNGOMLYFDSDTGALSTSTYSFSGTTLNVDVF 224

DB 172 PNVKQIDGKYYIGSDGQPKKNFALTNNKVLVFNKNTGALTDTTSQYQKQLTKLNNY 231

QY 225 SSHNKAYDSTAKSFELVNGYITANSWYRPAGILRNQGTWEASNENDLRPLVLSWMPDKOT 284

DB 232 TPNQIVNFNTSLKETIDNYVTADSWYRPKDKILKNGKWTASSESDLRPLLSWMPDKOT 291

QY 295 QVAYVYNNKYLSANETETVNTSETSVQDLNKEAQSTQTKIEQKITSNDSNTOWLRTAMEARV 344

DB 292 QIAYLNTNMQOGLGTGENYTTADSSQESLNLAAQTQVVKIETKISQTSQOTQOTQLRDIINSFV 351

```
QY 345 AAQPKWNMSTE--NPNKGDLHOGGALLYTNSDLTPWANSDYRLNRTPTQDGTCKKFT 401
Db 352 KTOPNWSQTESDTSAGESKXHLQGGALLYNSDKTAYANSDYRLNRTPTSGPKPYE 411
QY 402 EGEGGYEFLLSNDVNSNPVVAEQNLQHLNMGWDIVMGDKDANFGDVRVADVNVN 461
Db 412 DNSSGGYDFLLANDINDNSNPVVAEQNLQHLNMGYIVANDPEANFGDVRVADVNVN 471
QY 462 ADLLQVYSNYFKDNYKVTDSEANALAHISILEAWSLNDNQYNEDTNGTALSINDSRLTS 521
Db 472 ADLLQIASDYLKAHYGVDSKSEKNAIHLISILEAWSNDNDPOYNKDKTGAQLPDKLRLSL 531
QY 522 LAVLTQPCQRDLSN-----LISESVKBERANDTAYGDTIPTYSFVRHDSVQ 571
Db 532 LVALTR-PLEK-DASKNKEIRSGLEFVITNSLN-NRSEAGKNSERMANVIFTRAHDSEVQ 588
QY 572 TVIAKIVKEDITNSDGYFTFLDQKDAFKIYNEDMAKVNKYTHYNIIPAAVALLLSNME 631
Db 589 TVIAKIIKAIQNPKTDLGFTFLDELQKAFKIYNEDMRQAKKYTQSNIPATYALMSNKD 648
QY 632 SVPRVYVYGDLYTDDGQYMAKKSPYYDAIATMLQGRYAYVSGQSEBEVHKVNGNQ----- 686
Db 649 SITRLYVYGDWYSDGQYMATKSPYYDAITLLKARIKYAAGQDMKITVYEGDKSHMDW 708
QY 687 ---ILSSVRYGDLMSADDTQGTDLRSRTSGLVTLVNSNDPNLDLG-GDSLTVNMGRAHQ 742
Db 709 YTGVLTVRYGTCANBATO-QGSEATKTQGMVITNSNPSLKNQNDKVINMGTAHRNQ 767
QY 743 AYRPLILGKDGQVQSYLKSDS-NIVKYTDAMGNLTFTADDIKGYSTVDMSGYLAVWVPV 801
Db 768 EYRPLLLTTKDLGTSYSDAAKSLYRKTNDKGELVFDASDIQGLYNPQVSGYLAVWVPV 827
QY 802 GAKGDQVRVAADYNQKADGKSLKTSAAALDSQVYIEGFSNFQDFANNDADYTNKKIAENA 861
Db 828 GASDNQDVRVAASNKANATQVYESSALDSQLIYIEGFSNFQDFVTKDSYTNKKIAQNV 887
QY 862 DFEKLGITSPFMAQPVYVATGCSFLDSIIQNGYAFSDRYDLAMSKNNKYGSKDDLANAL 921
Db 888 QLFKSGVGVTSFMAQPVYVSESDGSLDSIIQNGYAFEDRYDLAMSKNNKYGQQDMINAV 947
QY 922 KALHANGIOAIDWPDQIYQLPGERVVTAKTNSYGNFTFDAYINNALYATNTKSSGD 981
Db 948 KALHKSIGIOAIDWPDQIYQLPGERVVTATRVNDYGEYRKDSEIKNTLYAANTKNSGKD 1007
QY 982 YQAYGGAFLDELKAKYPMFTNMISTGKPIDPSTKIKQWPAKYNGTNVLGKGAGYVL 1041
Db 1008 YQAKYGGAFLSLAAKYPISFNRTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVYVL 1067
QY 1042 SDDATGKYFTVNEGDFLPASFTGDQNAKTGYDGTGMAYYSTSGNKAIVSFIYEG-CH 1100
Db 1068 XONASDKYFELKGNQTYLPKQMT-NKEASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKN 1126
QY 1101 YYYFDHGHMVTGSYKAEDGNDYVYFLPNGIQMRDAIYDOAGNSYVYGRITGLYKGDWY 1160
Db 1127 WYTFDNGHMVVG-LQHLNGEVQYFLSNGVQLRESFLENADSGKNYFGLHGRY-SNGY 1184
QY 1161 PFVDPNNAKTVPRYPDANNVMAIGYRNNYMGOTTYFDPENGFOAKGOLLTDXG-THYPDE 1219
Db 1185 SF-----DNDSKWRYPDASGVMAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDD 1239
QY 1220 DNGMAKKNFVN-VGDDWYMDGNGNAVKGOYFVNNQILYFPNPTGVQVKGQFITDAQOR 1278
Db 1240 GSGNMAVNRFANDKNGDWTYLSNGDIALGVQVQTINGKTYFG-QDGKQIKGKIITD-NGK 1297
QY 1279 TSYDANSALKSSGFFTPNGSDWYVYAEYVYKGFQVAENQDQVYDQTTGKQAKGA 1338
Db 1298 LKXFLANSSELARNITAT-----DSQNNWYFG-SDGVAVTCG 1334
QY 1339 AKVDGRDLYFNPDSGVQVKGDFATDESNTSFPYHGDNGDKVYGGFFTTGNNAWYADNNG 1398
Db 1335 QTIAGKKLYFASD-GRQVKGVSFTV-YNGKVHYTHADSGELQVNRFEADKDGNNWYLDNSG 1392
QY 1399 NLVKGQFQIDGRWYHFEVDTGQAKGAALVNGQOLYFDVDSGIVQVKGDFVTDQGNYSY 1458
```

```
Db 1393 EALTGSQRI-----NGQRVFF-TREGKQVKGDVAYDERGLLRY 1430
QY 1459 DVNSGDKKVGCFPTTGDNAWYADQGKLAGRK 1492
Db 1431 DKNSGNMVKVVT-----LANGRR 1450

RESULT 9
Q59983_98TRE PRELIMINARY; PRT; 1590 AA.
AC Q59983_98TRE
DT 01-NOV-1996 (TREMREL. 01, Created)
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
DE 01-NAR-2004 (TREMREL. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfI;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OM2176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN (2)
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hetta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; 1HCX.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . ; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 Glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 42.5%; Score 3502.5; DB 2; Length 1590;
Best Local Similarity 45.4%; Pred. No. 2.1e-144;
Matches 749; Conservative 252; Mismatches 482; Indels 167; Gaps 40;

QY 1 MEKLLHYKLKHKGHWVTIAVASIG-LVSLVGAGTVSABDKVANDTTAAQTVDVTGQDQ 59
Db 1 MEKNVRFKMKHKVKKRWVTLVSASATMLASALCASVASADTDTASDDSNQAVV---TG-DQ 56
QY 60 ATTNDA--NTWTTDTDTADQSNQDQSGSNQDQAKODTANTDRNQADNSOTDNN 117
Db 57 TTNNQATQDTSIAATATSEQSASTD---AATDQASAAEQTQGTASTD--TAAQTTFNAN 111
QY 118 QATDQATSPATDG-TSVQRRDAANVATADQSGQATPSQEKSAALSLDNKLIDGKYYY 176
Db 112 EAKVPTENENQGFDEMLAEAKNVATA---ESDSIPSLAK-----MSNVKQVQDGKYYY 163
QY 177 VQADGSYKQNFATYNGQMLYFDSITGALSSSTSTYSFSGTITNLVDD---FSSHNKAYDS 233
Db 164 YDQDGNVKNFAVSGDKIYYFD-ETGAYKDTSKVDADKSSSAVSQNAITFAANNRAYST 222
QY 234 TAKSFELVNGYLTANSWYRPAGILNGQTWEASNENDLRPVLMSWMPDKDQTVAYVYWN 293
```

[illegible]

Db 1280 DKAGNWFYLGK-----DGAAVTGAOTIRGQKLYFKA-NGQV 1311

QY 1357 KGPATDESNGTSPYHGNDKGVGVGFFTTGNNAWYADNNG-----NLVKG----- 1403

Db 1316 KGDIVKTGDKIRYDAKSGBEQVFNKTVKAADGKTYVINGDVAVDPSPVVKGTQFXDASG 1375

QY 1404 ---PQETDGK-----WY-----HFDEV---TGQAQKGAALVNGQOLYFDVDSGIGQVKGDFV 1448

Db 1376 ALRFYNLKGQLVGTSGWYETANHDWYVIOGSKALTGEQTINGQHLVFKED-GHQVKGQLV 1434

QY 1449 TDGGGNTSYDDVNSGDKKVGFFTTGDNWY-----DGAAGV----- 1480

Db 1435 TGTGKRYRYDAMSGDQAFNKSVTVNGKTYFGNDGTAQTAGNPKGQTFKDGSDIRPYSM 1494

QY 1481 -----ADGO-----GNLAKGRKSIDNODLYPDPAATGKQVKGQLV-SIDGRN 1520

Db 1495 EGQLVGTSGWYENAQGWLVVKNKGLVTGLQTVGSQRVYFD-ENGLOAKGKAVRTSDGKI 1553

QY 1521 YPDSGSGNAKORFVRIGDQWLYFGNDGA 1550

Db 1554 RYFDENSGSMITNQWKFVYGQYVYFGNDGA 1583

RESULT 10

GTFF_STRMU STANDARD; PRT; 1476 AA.

AC P08987; O69381; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfb; OrderedLocusNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5.
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfb gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=95701124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22255063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., Mcshan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -1- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha

QY 765 NIVKYTDANGLTFTADDIKGYSTVDMGSLAVWVPVGVAKDQDVRVAADTNOKADGKSL 824
 Db 760 GLVRYTNDRGELFTTADIKGYANPQVSGYLVGVAAADQDVRVAASTPSTDKGSV 819
 QY 825 KTSAAALDSOVIYEGFNFODFANNDADYTNKKIAENADPFKGLGITSFEMAPQYVSATDG 884
 Db 820 HQNALDSRVWFEGFNFQAFATKBEYTNVJAKNDVDFAEAGVTDFFEMAPQYVSSTDG 879
 QY 885 SFILDSIQNGYAFSDRYDLAMSNNKYGKDDILANALKALHANGIOAIAADWPDDQIYOLP 944
 Db 880 SFILDSVIQNGYAFSDRYDLGISKPNKYGTADLLVKAIKALHSGIKVMADWPDMQYAP 939
 QY 945 GEEVWTAKTNSVGNFTFDAYINNALYATNTKSSGSDYQAYCGAFDELKAKYPMFTV 1004
 Db 940 EKEVWTATRVDFKFGKVEGSGIKSVLYVADSSGSDGQAKYGAFLLELOAKYPELFA 999
 QY 1005 NMISTCKPDPSPKIKQWAKYFNGTNVLGKAGYVLSDDATGKYFTVNEGDG--FLPAS 1062
 Db 1000 KQISTGVNPDSPVKIKQWAKYFNGTNILORGAGYVLDKQATNTYFNISDNKEINFLPK 1059
 QY 1063 FTGDQNAKTGFYDGTGMAYYSTSGNKAVNSFIYEGGHHYFDKOGHMYTGSYKABDGD 1122
 Db 1060 LL-NODSQVGSYDGGKGVYVYSTSGYQAKNTFISEGDKWYFDNNGYMTGA-QSINGV 1117
 QY 1123 YYPELNGIQWRDAIYQDAQNSYVYGRGILYKGDWYFPVDPNNAKTVFVRYFDANNVM 1182
 Db 1118 YYPELNGILQRLDAILKNEBDGTVAYVYNGDGRRYE-NGYQFM-----SGVWRHFN-NGEM 1169
 QY 1183 AIGYRMYGTYVYFDENGFOAGKOLLTDDKG-THYFEDENGAMAKNFV-NVGDOWYND 1240
 Db 1170 SVGLTWIDQVQVYFDENGFOAGKFTVTDGKIRYFDKSGNNYRNFLENBGRWLYLG 1229
 QY 1241 GNGNAVKGQPVNNQILYFNPETGVQVKGQFIDTAQRTSYDANSALKSSGFFTPNGS 1300
 Db 1230 EDGAATVGSQTNGQHLHYF-ANGVQVKGFEVTDYGRISYDSDSGDQIRNRF----- 1282
 QY 1301 DWYBAENGVIYVGFQVAENQDQYVYFDTTQKQAKGAUKVGRDLVFNPDGQVKGDF 1360
 Db 1283 -----VRNAQGWYFED-NGYAVTGARTINGQHLHYF-NGVQVKGFE 1324
 QY 1361 ATDESGNTSFYHGDNGDKVVGFFTTGNNAWYADNNGNLVKGFEIDGKWHFDEVTQ 1420
 Db 1325 VTRHGRISYDSDSGDQIRNRFVNAQGWYFONNGYAV----- 1365
 QY 1421 QAKGAALVNGQQLYFVDSGQVKGDFVTDGQNTSYDVNSGDKKVGFFTTGDNWY 1480
 Db 1366 --TGARTINGQHLHYF-NGVQVKGFEVTDYGRISYDSDSGDQIRNRFVNAQGWY 1422
 QY 1481 ADGQGNLAKGRKSIDNQDLYFDPATQVKGQLVSD-GRNYFDFSGSGNMAK 1532
 Db 1423 FDNNGYAVTGARTINGQHLHYF-RANGVQVKGFEVTDYGRISYDSDSGSERVR 1474

RESULT 11

Q55263 9STRE PRELIMINARY; PRT; 1590 AA.
 AC Q55263;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE GTF-I.
 GN Name=Glucosyltransferase;
 OS Streptococcus sobrinus.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1310;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC 33478;
 RA Sato S;
 RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase
 RT produced from Streptococcus sobrinus ATCC 33478."
 RL Kagoeshima Daigaku Shigakubu Kiyo 16:23-29(1996).

[2]

PROTEIN SEQUENCE.
 MEDLINE=91224988; PubMed=1827439;
 RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
 RT "Isolation and sequence of an active-site peptide containing a
 RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
 RT glucosyltransferases."
 RL J. Biol. Chem. 266:8916-8922(1991).
 DR EMBL; D63570; BAA09792.1; -, Genomic_DNA.
 DR PIR; A39841; A39841.
 DR HSSP; P06653; 1GVW.
 DR CO; GO:0003250; P:glucan biosynthesis; IEA.
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 3.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;
 Query Match 42.3%; Score 3481.5; DB 2; Length 1590;
 Best Local Similarity 45.3%; Pred. No. 1.8e-143;
 Matches 748; Conservative 249; Mismatches 487; Indels 167; Gaps 41;
 QY 1 MEKKLHYKLHKYKHHVITAVASIG-LVSLVGAGTVSAEDKVANDTTAQAATVGVGDQGO 59
 Db 1 MEKNVRFKHKYKRWVTLVSASATMLASALGASVASADTDITASDDSNQAVV---TG-DQ 56
 QY 60 ATTNDAA--NTNTTDTDTAQSNNTQDQAGSQNNQDQAKQDQTANTDRNQADNSOTDNN 117
 Db 57 TTNQATQDTSAAATATSEQASSTD---AATQASAAEQTQGTASTD--TAAQTITNAN 111
 QY 118 QATDQATSPATDG-TSVRRDAAAVATAADQEGQATPAPSEKSAALSNDVKLIDGKYV 176
 Db 112 EAKWVPTENENQGTDEMLAEAKNVATA---ESDFFPSDLAK-----MSNVKQVDGKYV 163
 QY 177 VQADSYKKNFAYITNGQMLYFDSGTGALSSTSTYSFSGITNLVDD---FSSHNKAYDS 233
 Db 164 YDQDGNVKNKFAVSGDKIYIFD-ETGAYKDTSKVDADKSSAVSQNATIFAANNRAYST 222
 QY 234 TAKSPELVNGYLTANSWYRPAGILRNGOTWEASNENDLRPLVMSWMPDKDQTOVAVYVYN 293
 Db 223 SAENFEADVNLITADSWYRPKILKDGKTWETSGKDDFPPLMANWPDTEKRYVYVYN 282
 QY 294 KYLSANETEVTNETSQVDLNKEAQSIQTKIEQKITSDNSTOWLRTAMEAFVAAQPKWNMS 353
 Db 283 LVVGIDKT-YTAETSQADLTAALVQARIEQKITTEQNTKWLREASIAFAVTKPQMWGE 341
 QY 354 TENFKGHLQGGALLYTN-SDLTPWANSDYRLLRTPQDGT-KKYPTGEGE---GGY 408
 Db 342 SEK-PYDDHLQNGALKFDNQSDLTPTQTSNYRLNRTPNTQTSGLDSRFTYNANDPLGGY 400
 QY 409 EFLLSNDVNSNPVQAEQLNQLHYLMWGDIVMGDKDANFDGVVDAVDNVNADILQVY 468
 Db 401 EFLLANDVNSNPVQAEQLNQLHYLMWGDIVMGDKDANFDGVVDAVDNVNADILQVY 460
 QY 469 SNYPKNTKYVTDSEANALAHISILEAWSLNDNQYNEGTNGTALSIDNSRLTSLAVLTQ 528
 Db 461 SDYLKAAVYGDKNKNNKANNHVSIVEASDNDTPYLHDDGDNLMNMDNFKRLSMLSLAKP 520
 QY 529 PCQRIDLNLNLSSESVNKERANDTAYGDTIPYTSFVRAHDSSEVQTVIAKIVKSKIDTNSDG 588
 Db 521 LDKRSGLNPLIHNSL-VDREYDDREVETVPSYFARAHDSVEQDIIRDIKAEINPNSFG 579
 QY 589 YTFLLDQLKDAFKIYNEDMAKNKTYTHYNTIPAAVALLSNMESVPRVYVGYDLYDDGGY 648
 Db 580 YSFQEEEDQAFKIYNEDLKKTKYTHYNVPLSYLLLTNKGSIPTVYIGDMFTDDGGY 639
 QY 649 MAKSPYYDAIATMLQGRYATVSGQSEVHKVNGNNOILSSVRYGQDLMSADDTQGTDL 708
 Db 640 MANKTVNYDAIESLLKARKMYVSGGQAMQNYQI-NGEILTSVRYGKALKQSD-KGDAT 697
 QY 709 SRTSGLVTLVNDPNLDLGGDSLTVMGRANQAYRPLILGTQGVQSYLKDSDTN--- 765
 Db 698 TRTSGVGMGNQPNFSLDGKVVALLNMGAHANQBYRALVSTKQGVATYATDADASKAG 757

```
QY 766 IVKYTDANGNLFTADDIKGYSTVDMGVLAVWVPVGAQGDQDVRVAADTNQADCKSLK 825
DB 758 LVKRTDENGVLFLNDLGLGVANPQVSGFLQVWVPVGAADDQDIRVAASDTASTDGKSLH 817
QY 826 TSAALDSQVYIEGFSNFQDANNADYTNKKIAENADFFKKLGITISFEMAPQVVSATDGS 885
DB 818 QDAMDSDRVWPEGFSNFQSFATKEEYTNVWANNVDKFKVSGITDFFEMAPQVVSSTDOQ 877
QY 886 FLDSIIQNGVAFSDRYDLAWSKNNKYGSKDDLANALKALHANGIQAIADVPDQYQLPG 945
DB 878 FLDSVIQNGVAFSDRYDLGSKANKYGTADQLVKAIKALHAKGLKVMADWVDPQMYTFPK 937
QY 946 EEWVTAKTNSVGNPTFDAYINNALYATNTKSSGSDYQAOYGAGLDELKAKYVDPMTYN 1005
DB 938 QEVVTVTRTDKFGKITAGSQINSHSLYVTDTKSGDDYQAKYGAFLDELKERYPELFTKK 997
QY 1006 MISTGKPIDPSTKIKOWEAKYFNGTGNVLGKAGYVLSDDATGKYFTVNENGBDFLPASFTG 1065
DB 998 QISTGQAIQPSVKIKQWAKYFNGSNILRGADYVLSQVSNKYFNVASDITLFLPSSLLG 1057
QY 1066 DONAKTFYDGTGMAYYST-SGNKAVNSFIYEGGHYYFDKXGHWMTGYSYKAEDGNDYY 1124
DB 1058 -KVESGIRYDGGKGIYNSSATGQVQKASFI TEAGNLYYFGKGYVMTGA-QTINGANYF 1115
QY 1125 FLPLNGIOMDQAIYQDAQNSYYGRTGILYK-----GDNWYFVDPNNANKTVRYFPA 1178
DB 1116 FLENGTALRNTIYTDAGNSHYANDGKRYNGYQQFGNDW-----RYFKD 1161
QY 1179 NNMAIGYRNMVYQOTYYFBNQGAQKQLL-TDDKGYHYPDEBNGAMAKNFV-NVGDDW 1236
DB 1162 GN-MAVGLTVDGNVQYFKDGQVQAKDKIIVTRDGKVRFPDQNGNAVNTFIADKTHW 1220
QY 1237 YMDGNGNAVQGOYPVNNQILYNPETGVQVKGQFITDAQRTSYVDANSALKSSGFFT 1296
DB 1221 YVLGKDGVAVTGAQTVGKQKLYPE-ANGQVKGDFVTSHEGKLYFYVDVSDGDMWTDFT 1279
QY 1297 PNGSDHYAENGYYVYKGFQVAENQDQWYFDOTTCKQAAGAAGVDRGLYFNPDGVOY 1356
DB 1280 DRAGNMFYLGK-----DGAAYSGAQTIRGQKLYFKA-YGQOV 1315
QY 1357 KGDFAATDESNTSFYHGDNGDKVGVGFFTTGNNAWYADNNG-----NLVKG----- 1403
DB 1316 KGDIVKGTOKIRYDAKSGEQVFNKTVKAADGKTVIGNGVAVDPSVVKGQTFKDSG 1375
QY 1404 ---FQIBDGK-----WY-----HFDEV---TGQQAAGALVNGQQLYFDVDSGIVKGDV 1448
DB 1376 ALRFYNLQQLVTGSGWYETANHDWYIYQSGKALTGEOQHLYFKED-GHQVKGQLV 1434
QY 1449 TDQGNNTSYVDNSGDKKNGFFTTGDNWY----- 1480
DB 1435 TRTDGKRYRYDANSBGDAFNKSYTVNGKTYFYFNGDGTAGTGNPKQIIFKDGSLVLPYSM 1494
QY 1481 -----ADQO-----GNLAKRKSITDNDLYFDPATGQVKGQLV-SIDGRN 1520
DB 1495 EGQVLVSGSWYNAQQQWLKYVKGKVLTLGLQTVGSRVYFD-ENGIOAKGKAVRTSDGKI 1553
QY 1521 YPDSGSGNMAKNRFRVIRIGDQWYIFGNDGAA 1551
DB 1554 RYFDENSGSMITNQWKFVYQYQYFYFNGDGA 1584

RESULT 12
ID GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-1) (Dextranucrase)
DE (sucrose 6-glucosyltransferase).
GN Name=gTfI;
OS Streptococcus downei (Streptococcus sobrinus).
```

```
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RC MEDLINE=87308014; PubMed=3040686;
RX Ferreretti J.J., Gilpin M.L., Russell R.R.B.;
RA "Nucleotide sequence of a Glucosyltransferase gene from Streptococcus
RT sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278 (1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6-linkages), GTF-S synthesizes both
CC water-soluble glucans (alpha 1,6-linkages), GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 19 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M17391; AAC63063.1; -; Genomic DNA.
DR InterPro; IPR002479; Cell_wall_bd_put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1..38 Potential.
FT CHAIN 39..1597 Glucosyltransferase-I.
FT REPEAT 1099..1132 A repeat.
FT REPEAT 1163..1213 AC repeat.
FT REPEAT 1227..1277 AC repeat.
FT REPEAT 1292..1342 AC repeat.
FT REPEAT 1352..1399 B repeat.
FT REPEAT 1406..1455 AC repeat.
FT REPEAT 1465..1512 B repeat.
FT REPEAT 1519..1568 AC repeat.
FT REPEAT 1582..1597 A repeat (incomplete).
FT REGION 39..1050 Catalytic (approximate).
FT REGION 1099..1597 1.25 A, 2 B and 5 AC repeats.
FT REGION 1099..1597 Glucan-binding (approximate).
FT SEQUENCE 1597 AA; 177080 MW; B9E86A200868798B CRC64;
SQ
Query Match 42.3%; Score 3481; DB 1; Length 1597;
Best Local Similarity 44.6%; Pred. No. 1,9e-143;
Matches 739; Conservative 261; Mismatches 484; Indels 174; Gaps 40;
QY 1 MEKLLHYKLHKYKHWTVIAVASIG-LVSLVGAGTVSAEDKVANDTTAQATVGVDTGDDQ 59
DB 1 MEKNERFKMHKVKRWVTVISVASATMLASALGASVASADTETVSDSNQAVLTA----DQ 56
QY 60 ATTND--ANTNTTDTDTADQSANTN--QDQA--GSDSNQDQAQKODTANTDRNQADNSOT 114
DB 57 TTTNQDTQETSVAATATSEQSASTDAATDQASAAEQTQGTASTD--TAAQTFTT 114
QY 115 DNNQA----TDQATSPATDGTSVQRDAANVAATAADQEQATAPSEQEKSAALSLDNVKLI 170
DB 115 NANEAKWVPTENENQVFTD---EMLAEXKNVATA---ESNSIPSLAK-----MSNVKQV 163
QY 171 DGKYYVQADGSYKKNFPAITVNGQMLYFDSDFTGALSSTSYFSQGTNNLVDD---FSSH 227
DB 164 DGKYYVQDQGNVKNKFAVSGEKIYFVD-ETGAYKDTSKVEADKSGSDISKEETTTFAAN 222
```


KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.	
FT SIGNAL	1 38 Potential.
FT CHAIN	39 1592
FT REPEAT	1093 1142
FT REPEAT	1158 1207
FT REPEAT	1222 1272
FT REPEAT	1287 1337
FT REPEAT	1402 1451
FT REPEAT	1514 1563
FT REPEAT	1577 1592
FT REGION	39 1044
FT REGION	1093 1592
FT REGION	1093 1592
SQ SEQUENCE	1592 AA; 176168 MW; BC0A66D079351ECF CRC64;
Query Match	
Best Local Similarity 42.2%; Score 3472.5; DB 1; Length 1592;	
Matches 741; Conservative 254; Mismatches 496; Indels 153; Gaps 39;	
QY	1 MEKKLHYKLVKVKHWTIAVASIG-LVSLVGAGTVSAEDKVKANDTTAAQTAVGVDTGDDQ 59
Db	1 MEKNVRFKMHKVKGRWVLSVASATMLASALGASVASADTDTASDDSNQTVV---TG-DQ 56
QY	60 ATTNDAA--NTNTTDTDDQSANTNODQAGSDQSNNOQAKODTANTDRNQADNSQTDNN 117
Db	57 TTNNQATDQTSIAATATSEQSASTD---AATDQASAAEQTQGTASTD---TAAQTITNAN 111
QY	118 QATDOATSPATDG-TSVQRDRDAANVATAADQBGOTAPSEQKSAALSNDNVKLIDGKY 176
Db	112 EAKWPTENEGPTDEMLAEAKNVATA---ESDSIPSLAK-----MSNVKQVDGKY 163
QY	177 VOADSGYKKNFALTIVNGQMLYFSDTGLSSTSTYSFSGGTTNLVDD---PSSHKA YDS 233
Db	164 YDQDGNVKNFPAVSVDGIYYFD-ETGAYKDTSKVDADKSSSAVSQNAITFAANNRAYST 222
QY	234 TAKSFELVNGYLTANSWRPAGILRNGOTWEASNENDLRPVLMSWPKDQTOVAVNYMN 293
Db	223 SAKNFEADVNLITADSWRPKSLIKDKGTWTSKGDDFRPLLMWMPDPTETKRNVTNN 282
QY	294 KYLSANETEVNTEVSODILNKEAQSITQKIEKITSNDSNLTWLRAMEAFVAAQPKWMS 353
Db	283 KVVGDIKT-YTAETSOADLTAAELVQARIEQKITSNNKWLREASIAFVKTQPMNGE 341
QY	354 TTFNKGDLHOGGALLYTN-SOLTFWANSDYRLNRTPTQDQGTGK-----KYTEGEGGY 408
Db	342 SEK-PYDDHLQNGALLFDNQDLDTPDQSNRYLLNRTPTNQTGSLDSRFTYFNPDPFGY 400
QY	409 ELLSNDVDNSPVVQAEOLNOLHYLMWGDIVMGDKANPDGVRDADVNNADLLQVY 468
Db	401 DFLANDVDNSPVVQAEOLNOLHYLMWGDIVMGDKANPDGVRDADVNNADLLQVY 460
QY	469 SNYFKDNYKVTDSEANALAHISILEAWSLNDQYNETDNGTALSIDNSSRLTSLAVLTQ 528
Db	461 SDYLKAAVGIKNNKANNHVSIVEASNDPTFLHDDGDNLMNDNKFRLSMLWSLAKP 520
QY	529 PQGRIDLNLISESVNKEPANTAYGDTIPTYSFVRAHDSVQTVIAKIVKEKIDTNSDG 588
Db	521 TDVRSGLNPLIHNSL-VDREVDDREVETVPSFARAHDSVQDIIIRDIKAEINPNSFG 579
QY	589 YTFLLDOLKDAKPIYNEDMAKNTYTHYNI PAAYALLSNMESVPRVYVGYLTDGQY 648
Db	580 YSFTQEEIDQAFKIYNEDLUKSKDKYTHYNVPLSYTLTLTNKSGISPRVYVGYDMFTDGGY 639
QY	649 MAKSPYDAIATMLQGRITAYVSGQSEEVHVKVNGNQNILSSVRYGQDILMSADDTQGTDL 708
Db	640 MANKTVNYDAIESLLKARKMYVAGQAMQNYQI-NGEILTTSVRYGKALKQSD-KGDAT 697
QY	709 SRTSGVLTVLSNDPNLDLGGDSLTVNMGRANQAYRPLILGTCKGVQVSLKDSQTN--- 765
Db	698 TRTSGVGVVGNQPNFSLDGLKVVALMGAHAHQBYRALMVSTKDGCVATYATDADASKAG 757
QY	766 IVKYTIDANGLFTADDIKGYSTVDMSGVLAWVPVGAKGODVRAADTNQADKGSILK 825

Db	758 LVKRTDENGYLFLNDDLKGVANPQVSGFLQVWPVGAADDQDIRVAASDTASTDGKSLH 817
QY	826 TSAALDSQVIYEGFSNFQDPANNDADYTNKKIAENADFFKKLGLITSFEMAPQYVSATDGS 885
Db	818 QDAAMDSTRMFEFGFSNFQSPATKEEYTNVVIANNVDKPFVSGWIGITDFEFMAPOYVSTDGQ 877
QY	886 FLDSITQNGYAFSDRYDLAMSKNNKYGSKDDLALNALKALHANGIOAIADWVPDQIYQLPG 945
Db	878 FQDSVIQNGYAFTRDRIYLGMSKANKYGTADQLVKAIKALHAKGLKLMADWVPDQMYTPK 937
QY	946 BEVTTAKTNSYGNPTFDAYINNALYATNTKSSGSDYQAOYGGAFDLBILKAKYPMFTVN 1005
Db	938 QEVTVTRTDFEGRPIAGSQIHSLIYVTDTKSSGDDYQAKYGAFLDELKEXYPELFTCK 997
QY	1006 MISTGKPIDSTKIKQWEAKYFNGTNGVLGKAGYVLSDDATGKYTYTVNENGDFLPSFTG 1065
Db	998 QMSTQQAIDPSPVKIKQMSAKYFNGSNILGRGADYVLSQVSNKYFNVASDTLFLPSLLIG 1057
QY	1066 DONAKTGFYDGTGWAYYST-SGNKAVNSFYEGGHYVYEDKDHMTVSGYKAEQNDYV 1124
Db	1058 -KVESGIRYDGGKYIYNSSATGQVKASFITEAGNLYYFGKDGVMVTGA-QTINGANF 1115
QY	1125 FLPNGIQWRDAIYODAOQNSYYYGRGTILYKGDNNYFPVDPNNANKTVFRYFDANNVMAI 1184
Db	1116 FLENGTALRNTIYTDACNSHYANDKRYENENGYQQFGND-----WRYFKDGN-MAV 1168
QY	1185 GYRNNYGYTYFDENGFOAKQOLL-TDDKGYTHYFDENGAMAKNFV-NVGDDWYMDGN 1242
Db	1169 GLTVDGNVQYFDKQVQAKDIIIVTRDGRVRYFDPHNGNAVNTFIADKTHWYLYLGD 1228
QY	1243 GNAVKGYPVNNOILYFNPETGVQVQGFITDAGRTSYVDANSALKSSGFFTPNGSDW 1302
Db	1229 GVAVTGAQTVGKQKLYFE-ANGQQVKGDFVTSDEGKLYFYDQVDSGDMWTDTFIEDKAGNW 1287
QY	1303 YYAENGYVYKGFQVAENQDQWYFDQTTGQAKGAAGVDGRDLYFNPDSPGVQVKGDFAT 1362
Db	1288 FYLKG-----DGAAVTGAQTIQKLYFKA-NGQQVKGDIVK 1323
QY	1363 DESGNTSPYHNGDKVVGSGFTTGNNAWYVADNNG-----NLVKG-----PQ 1406
Db	1324 GTDQKIRYDAKSGEQVENKTKAADGKTYVIGNDGVAVDPSVVRKGQTFKQASGALRFYN 1383
QY	1407 IDGK-----WY-----HFDEV---TCQAKGAALVNGQOLYFNDVDSGIOVKGDFVTDGQ 1454
Db	1384 LKGQVLTSGWYETANHDWVYIQSGKALTGEGTINGQHLHYFKD-GHQVKQLVGTGDK 1442
QY	1455 TSYDVNSGDKKVGFFTTGDNWY-----PRT; 1577 AA. 1480
Db	1443 VRYTDANSQDQAFNKSVTNGKTYFYFGNDGTAQTAGNPKGQTFKDGSDIRFYSMEGQLVT 1502
QY	1481 -----ADGQ-----GNLAKRKSIDNQLYFDPATGKQVKGQLV-SIDGRNYYFDSG 1526
Db	1503 GSGWYSNAQGWLYVKNKVLTLGLTVGSRVYFD-ENGIOAKGAVRTSDGKIRYFDEN 1561
QY	1527 SGNAKQPFVRIQDQWYIFGNDGA 1550
Db	1562 SSGMITNQWKEVNGRYFYFGNDGA 1585
RESULT 14	
Q55265 STRSL PRELIMINARY; PRT; 1577 AA.	
AC	Q55265;
DT	01-NOV-1996 (TremBLrel. 01, Created)
DT	01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT	01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE	Glucosyltransferase precursor.
GN	Name=gtfm;
OS	Streptococcus salivarius.
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=1304;
RN	[1]

RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL: L35928; AAC41413.1; -; Genomic_DNA.
DR PIR: T30858; T30858.
DR HSSP: P06653; 1H8G.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro: IPR004829; C:aurface antigen.
DR InterPro: IPR002479; C:binding.
DR InterPro: IPR003318; Glyco_hydro_70.
DR Pfam: PF01473; C:binding 1; 4.
DR Pfam: PF02324; Glyco_hydro_70; 1.
DR ProDom: PD153432; C:aurface_antigen; 1.
KW Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1577 Glucosyltransferase.
SQ SEQUENCE 1577 AA; 175290 MW; 3EFB988A7D3A7BF3 CRC64;

Query Match 41.6%; Score 3429.5; DB 2; Length 1577;
Best Local Similarity 44.7%; Pred. No. 3.2e-141;
Matches 738; Conservative 250; Mismatches 471; Indels 193; Gaps 40;

QY 1 MEKHLKHLKVKHVVITIAVASIGLSVLGAGTVAEDKVAANDTTA----- 47
DB 1 MENKVRFLKHLKVKHVVITIGVITLSMVALAG-GSLLAQKVEADETSAPNGDGLQQLSED 59
QY 48 -----QATVGVDGTQDQATNDANTTTDTADQSANTNPQAGSDQSNQDQAKQDT 101
DB 60 GTASLVTITVTTEQSAQASAVATASVSHETSFOAATSQVSAQTAQOTSPVASQEV 119
QY 102 ANTRDQADNSQD-----NNQATDQATSPATDGTSPQRD-----AANVAT-A 144
DB 120 AVSSQTSQSGQETQTEQVSQGTSTQVAGQTSQASTPSVTEQARPRVLTAAPAIAATRA 179
QY 145 AD-----QEGQ- 150
DB 180 ADSTIRANRNTNITITAGTTPNTVITIGTNPENVTVPSPNGTRPNVITVQNPQ 239
QY 151 -----TAPSEQKSA-----NKLIDGKYVVOAGSYYKKNFAITVNGQM 195
DB 240 NKEVPQSPQSPQNPQPSLDYKPVASNLKTIDGQYVVE-NGVVKKNAIAELDGL 298
QY 196 LYFDSDTGAL--SSSTSYFSQGTTLNVDVDFSHNKAYDSTAKSPFLVNGYLTANSWRP 253
DB 299 YTFD-ETGAWVDQSKPLRADAI PNNSI--YAVYNAQYDTSKSPFHLNDFLTADSWVRP 355
QY 254 AGILRNGQTEASNDLRPVLMSWMPDKDTQAVYVYNNKYL SANETEVTNETSOVDLN 313
DB 356 KQLKDGKNWTASTEKDRPLMTWMPDKVTQVNLNYSQQGFGNKYITTDMS-YDLA 414
QY 314 KEAQSIQTKIEOKITSNSTQWLTAMEAFVAAQPKNNMSTF-NFNKG-DHLQGGALLYT 371
DB 415 AAAETVQRIEGRICREGNTTWLRQLMSDFIKTQPCWNSSESDNLLVGKDLHGGALTF 474
QY 372 NSDLTPWANSYRLNRRPTQODGKTKYFTEGEGGYEFLLSNDVNSPVPVQAOLNQL 431
DB 475 NNSATSHANSDFRLNRRPTTQGTGRKYHIDRSNGGYELLANDINDNSPAPVQAOLNWL 534
QY 432 HYLNMWGDIVMGDKANFDGVRDADVNVDNADLLQVYSNYKDYNKVTDSEANALAHISI 491
DB 535 HYIMWIGILNDPSPANFDGVRIDAVDNDVADLLQIASDYFKEKTRVADNEANAIAHLSI 594
QY 492 LEAMSLNDQYNDTNGTALSDNSRLTSRLAVLTQPCQRIDLSNLISSESVKNERANDT 551
DB 595 LEAMSYNDHQYNDKTKGAQLSDNPLRETLTTLFLRKSNYRGSLSERVITNSLN-NRSEQ 653
QY 552 AYGDITPYFSVRAHDSQVTVIAKIVKEIDTNSDGTFTFLDQLDKAFKYNEMAKVN 611
DB 654 KHTPRDANYIFVRAHDSQVAVLANIISKQINPKPTDGTFTMDLQKQAFETYNADIADK 713

RESULT 15

GTFC_STRMU

ID GTFC_STRMU STANDARD; PRT: 1455 AA
AC P13470; O69382; O69385; O69386; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)

QY 612 KTYTHYNI PAAYALLSNMESPRVYVYGLYTDGQYMAKSPYDAIATMLOCRAYVS 671
DB 714 KKTYQNI PAAYATMLTKDSITRVYVYGLYTDGQYMAEKSPYDAIALLARIKYVA 773
QY 672 GGQSEVHKVNGNQNQLSSRVYVYGLYTDGQYMAEKSPYDAIALLARIKYVA 730
DB 774 GGQDMKVTKLNG-YEIMSSRVYVYGLYTDGQYMAEKSPYDAIALLARIKYVA 831
QY 731 LTVNMGRAHANQAYRLLILGTGQVOSYKLDSD--TNIYKYTDANGNLFTADDIKYST 788
DB 832 LVNMGRAHANQAYRLLILGTGQVOSYKLDSD--TNIYKYTDANGNLFTADDIKYST 891
QY 789 VMSGYLA VVPVPGAGKQDVRVAADTNQADKSLKTSAAALDSQVYIEGFSNFQFPANN 848
DB 892 VEVSGYLA VVPVPGASENDARTKASTKKGE-QVFESSAALDSQVYIEGFSNFQFPANN 950
QY 849 DADYTNKKIAENADPFKLGITSPFMAPOQVVSATDGSFLDSIIQNGYAFSDRYDLAMSKN 908
DB 951 PSQYTNRVIAQNAKLFKEWGITSPFMAPOQVVSATDGSFLDSIIQNGYAFSDRYDLAMSKN 1010
QY 909 NKYGSKDDLANALKAHLANGIOAIAADWVDPDIYOLPGEVVTAKRTNSYGNPTFDAYINN 968
DB 1011 NKYGSUKDLMDALRALHAEIGISAIAADWVDPDIYOLPGEVVTAKRTNSYGNPTFDAYINN 1070
QY 969 ALYATNTKSSGSDYQAQYGGAFDELKAKYPMFTVMNISTGKPIDPSTKIKOWEAKYFN 1028
DB 1071 SLYAAKTRTFGNDFOGKYGGAFDELKAKYPAIFERVOISNGRKLTKNEKITOWSAKYFN 1130
QY 1029 GTNVLKGGAGYVLSDDATGKYFTVNENGDFLPASFTGQNAKTFYFYDGTGMAVYSTSGN 1088
DB 1131 GSNIGTGAHYVLDNATNQYFSVAKQGTFLPKQMT--EITGSGFRVRVDDVYLSIGGY 1188
QY 1089 KAVNSFIYEGGH-YVYFQKDGHWYTSYKAEQNDYFPLNGIOMRDALYQDAQNSYYY 1147
DB 1189 LAKNTFIQVGAQWYFEDKNGNMVTGE-QVIDGKYFFLDNGLQLRHVLRQSGDGHVY 1247
QY 1148 GRGTGI-LYKGDNWPFPVDPNNAKTVRYFDANNVMAIGYRNMYGQTYVYFDE-NGFQAKG 1205
DB 1248 DPKGVQAFNG--FYDPAGP---RQDVRYFDGNGQMYRGLHDMYGTTFYFDEKGTGLOAKD 1301
QY 1206 QLLT-DDKGTHTYFDEBNGAMAKNFVNVGDD--WYMDGNGNAVKGQYVNNQILYFNPE 1262
DB 1302 KFIREDGRTRYFIPDTGNLAVNRFAQNPENKAWYLDNSGYAVTGLQTINGQYFYDNE 1361
QY 1263 TGVVQVGQFIITDAQRTSYDANSALXSSGFTTNGSDWYVAENGYYVYKGFQVAENQD 1322
DB 1362 -GRQVKGHFVT-INNQRYFLDGDGSEIAPSRFVTEN-----N 1396
QY 1323 QWYVYFDQTTGQAKGAAGVDRDLVFNPDGSGVQVKGDPATDESNTSFYHGDNGDKVVG 1382
DB 1397 KMYVYD-GNGKLVKGAQVINGHYVFNNDYS-QVKGAWA-----NGRYVDGSGQAVSNQ 1449
QY 1383 FPTTGNNAWYADNNGNLVKGFEIDGKWYHFDEVTGQQAAGALVNGQOOLYFDVDSGIG 1442
DB 1450 FQIAANQWYLVNQDGHKVTGLQINNNKYYVPG-----SNGAQ 1487
QY 1443 VKGDFVTQCGNTSYVDVNSGDKKYNNGFTTGDNNAWYADGQGNLAKGRKSTDNQDLYFD 1502
DB 1488 VKGKLLT-VQGGKCYFDAHTGEQVYVNRFEAARGCWYFNSAGQAVTGOOVINGKQLYFD 1546
QY 1503 PATGQVKGQVLSIDGRNYVYFDGSGSNMAKNR 1534
DB 1547 -GSGRQVKGRYVYVYGGKRLFCDAKTGELRQR 1577


```
QY 57 QOATNDANTWT-TDUTADQASANTNQOAG--SDQSNQDQAKQDTANT-----DRNQ 108
Db 56 ---VTTSEAAKELTATDSTATSATSQPTATVTDNVSTTQNTTANTANFDVKTPTT 112
QY 109 ADNSQTDNNQ---ATDOATS-----PATDGTSVQRREDAANVATAADQEG---OTAP 153
Db 113 SQSKTDNSDKIATSKAVNLTATGKFVANNNTAHSRTVTDKVIPIKPKGKQKQPS 172
QY 154 SQEKSAAI--SLDNVKLIDGKYVVOAGDSYKKNFAITVNGQWLYFDSOTGALSSTSTYS 212
Db 173 LSQDDIAALGNVKNIRKNGKYYYKEDGTQKNYALNINGKTFPD--ETGALSNNLTSP 231
QY 213 FSQGITN--LVDDPSHNKAYDSTAKSPELVNGYLTANSWYRPAGLRNGQWAEASNEND 270
Db 232 KKGNTFNNDNTNSFAQYNQVYSTDAANFEHVHLYTAESWYRKYILKDKGTWQSTEXD 291
QY 271 LRPVLMSWPDKDTQVAYVYNNKXLSANETEVNETSVDLNKEAQSOTKLEQKITSD 330
Db 292 FRPLMTWPPDETQRYVYNNQAQGIHQTYNT--ATSPQLNLAAQTIQTKIEKITAE 350
QY 331 NSTOMLRTAMEAFVAAQPKWNMSTEN--FNKGHLOGGALLYT--NSDLTPWANSDYRLNLR 388
Db 351 KNTNMLRQIISAFVKTQSAWNSDSEKFPD--DHLQKALLYSNNSKLTQANSNYRLNR 408
QY 389 TPTQODGTK--KYFTEGGGGGYFLLSNDVNSNPVQAEQNLQHLHLMWGDIVMGDXD 446
Db 409 TPTNQTGKDPRTADRTTGGYEFLLANDVNSNPVQAEQNLNLHFLNFGNIYANDPD 468
QY 447 ANFDGVRVDAVNVAADLQVYSNFYKNYKYTDSANALAHISTLEAMSLNDNQNEBT 506
Db 469 ANFDSIRVDAVNDAVADLQIAGDYLKAAGIKHNDKAANDHLSLEAWSYNDTPYLHDD 528
QY 507 NCTALSIDNSRLTSLAVITKPGQRI--DLNSLISESVNKERANDTAYGDTIPTYSFVRAH 566
Db 529 GDNMINMDNRLSLULSLAKPLNQBSGNPLITNSL--VNRDNDNAETAAVPSYSPIRAH 587
QY 567 DSEVQTVIAKIYKEIKDITNSDGYTFTLDQDKAFKIYNEDMAKVNKTYTHYNIPAAAYALL 626
Db 588 DSEVQDLIRIIRAEINPNVWGYSTFMEBEEKAFIYNKOLLATEKKYTHYNTALSYALL 647
QY 627 LSNMESVRVYVYDLYTDDGOYMAKSPYDYAIAITMLQRIAYVSGGQSEEVHKVNGNQ 686
Db 648 LTNKSSVRVYVYVYDLYTDDGOYMAKHTINYEAETLLKARIKYVSGQAMRNQV--GNSE 706
QY 687 ILSVRYGODLMSADTQGTDLTSRTSGLVTLVSNPDL--GQDLSLTVMGRAHANQAYR 745
Db 707 IITSVRYGKALKATDT--GDRTRISGVAVIENGPSPLEKASDRVVMNGAAHKNQAYR 765
QY 746 PILGTDGVSQSYLXDS--TNIYKYTDANGNITFTADDIKYSTVDMSCYLAVVVPVGA 804
Db 766 PLLLTDDNGIKAYHSDQEAAGLVRYTNDRGELIFTAADI--KGYANPQVSGYLGWVPVGA 825
QY 805 DQDVRVAADTNQKADGSKLTSALDSQVIYEGFSNFPDPAANDADYNTKKIAENADPF 864
Db 826 ADQDVRVAASTAPSTDGKSVHQNAALDSRMFEGFSNFOAFATKKEEYTNVVIKNDKF 885
QY 865 KKLGITSPEMAPQYVYATDGSFLDSTIONGYAFSDRYDLAMSKNNKYGSKDDLALNALKAL 924
Db 886 AEWGVTDEMAPQYVYSSDGSFLDSVIQNGYAFTRDYDLGISKPNKYGTDADDLVKAIKAL 945
QY 925 HANGIAIADWVPDQIYQIPGSEVVTAKRTNSYGNPTFDAYINNALYATNTKSSGSDYQA 984
Db 946 HSKGIKVMADWVPDQYALPEKEVVTATRVKYGTPVAGSQIKNTLYVVDGSSGKQQA 1005
QY 985 QYGGAFLELAKYKAPDMFTVMIMSTGKPIDPSTKI--KQWEAKYFNGTNVLGKAGYVLSDD 1044
Db 1006 KYGGAFLBELOAKYPELFARKQISTGVPMDPSVKIKQMSAKYFNGTNILGRGAGYVLDQ 1065
QY 1045 ATGKYFTVNEGDFLPASFT----GDQNAKTGFYVDGTGMAYYSTSGNKAVNSFIYEGGH 1100
Db 1066 ATNTYFSLVSONTFLPKSLVNPNNHGTSSSVTLVFDGKYVYVYSTSGNAKNAFISLGN 1125
QY 1101 YYYFDKGDHMTGSKYABDGDNDYFLPNGIQMRDAIYQDAQNSYVYGRGTILYK----- 1155
```

```
Db 1126 WYFDNNGYMTGA--OSINGANYIFLSNGITQLRNAIYDNGNKVLSYNGDREYENGYL 1184
QY 1156 -GDNWYFPVDPNNANKTVFRYFDANNVMAIGYRNMYGOTYYPDENGFOAKGOLLTDDKG- 1213
Db 1185 FQOQW-----RYFQ--NGIMAVGLTRIHGAVQVFPDASGFQAKGQFITTADGK 1229
QY 1214 THYFDEBDNGAMAKNFV--NVGDDWYMDGNGNAVKQYFVNNQILYFNPETGVQVKGQFI 1272
Db 1230 LRYFDRDSNQIISNRFPVRSKGEWPLFDHNGVAVTGTVTFNGQRLYFKP--NGVQAKGEFI 1288
QY 1273 TDAQRTSYDANSALKSSGFTPNGSDWYVAENGYVVKGFQKQVAENQDQWYFDQTTG 1332
Db 1289 RDADGHLRYDPSNGNEVRNF-----VZNSKGEWELFDH--NG 1325
QY 1333 KQAKGAAKVDRDLVFNPDGVOVKGDFAFDTSNGTSEYHGDNGDKVVGFFFTTGNAMY 1392
Db 1326 IAVTGRVNVGQRLYFK--SNGVQAKGELLITERKGRIKYIDPNSGNEVRNRYVRTSSGNWY 1384
QY 1393 YADNNGNLVKFQIEIDGKWHYHFDVETGQQAAGAAALVNGQQLYFDVDSGIVQKGDFTVD--G 1451
Db 1385 YFNGDGYALIGHVHVEGRVYFDE-----NGVYRYASHDQORNHWDYDVRDPG 1432
QY 1452 QGNTSYDVNSGDKXKNGVGFFTTGDN 1476
Db 1433 RGSASAVRFRHSR---NGFF---DN 1451
```

Search completed: February 11, 2006, 19:38:42
Job time : 222.688 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 135.868 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-39
Perfect score: 7230
Sequence: 1 MEKNRYKLHKVKQWVAIG.....WKNVDGNWYFNRGLATRW 1365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7230	100.0	1365	7	ADD93659 Streptococcus
2	7230	100.0	1365	9	ADK37282 Streptococcus
3	3174	43.9	1430	5	Aau98029 S. mutans
4	3174	43.9	1430	7	ADD93656 Streptococcus
5	3174	43.9	1430	9	ADK37279 Streptococcus
6	3169	43.8	1430	5	Aau98043 S. mutans
7	3168	43.8	1430	5	Aau98042 S. mutans
8	3168	43.8	1430	5	Aau98041 S. mutans
9	3163	43.7	1430	5	Aau98044 S. mutans
10	3163	43.7	1430	5	Aau98045 S. mutans
11	3080.5	42.6	1554	7	ADD93658 Streptococcus
12	3080.5	42.6	1554	9	ADK37281 Streptococcus
13	3031	41.9	1590	7	ADD93657 Streptococcus
14	3031	41.9	1590	9	ADK37280 Streptococcus
15	3026.5	41.9	1577	2	AAR91047 Alpha-D-g
16	3023	41.8	1592	2	AAR32925 Glucosyl
17	3001	41.5	1476	5	Aau979284 Streptococcus
18	2973.5	41.1	1475	5	Aau98038 S. mutans
19	2963.5	41.0	1475	5	Aau98037 S. mutans
20	2957.5	40.9	1475	5	Aau98036 S. mutans
21	2956.5	40.9	1475	5	Aau98035 S. mutans
22	2956.5	40.9	1475	5	Aau98034 S. mutans
23	2951.5	40.8	1475	5	Aau98033 S. mutans
24	2951.5	40.8	1475	5	Aau98032 S. mutans

25 2950.5 40.8 1475 5 AAU98031 S. mutans
26 2949.5 40.8 1475 5 AAU98040 S. mutans
27 2946.5 40.8 1475 5 AAU98030 S. mutans
28 2945.5 40.7 1475 5 AAU98027 S. mutans
29 2945.5 40.7 1475 7 ADD93654 Streptococcus
30 2945.5 40.7 1475 9 ADK37277 Streptococcus
31 2939 40.7 1499 7 ADC54806 Protein S
32 2929.5 40.5 1475 5 AAU98039 S. mutans
33 2893.5 40.0 1375 5 AAU98028 S. mutans
34 2893.5 40.0 1375 5 AAU79288 Streptococcus
35 2893.5 40.0 1375 7 ADD93655 Streptococcus
36 2893.5 40.0 1375 9 ADK37278 Streptococcus
37 2891.5 40.0 2835 5 ABR98574 Dextran s
38 2891.5 40.0 2835 6 ABR55594 Amino aci
39 2804 38.8 1518 7 ADD93660 Streptococcus
40 2804 38.8 1518 6 ADK37283 Streptococcus
41 2680.5 37.1 1497 6 ABR63234 Glucansuc
42 2661.5 36.8 1477 9 ADY72733 Mutant de
43 2656.5 36.7 1477 9 ADY72696 Mutant de
44 2653.5 36.7 1477 9 ADY72732 Mutant de
45 2640 36.5 1527 7 ADC54807 Leuconost

ALIGNMENTS

RESULT 1
ADD93659
ID ADD93659 standard; protein; 1365 AA.
XX
AC ADD93659;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus downei glucosyltransferase-S.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus downei.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
(FORS-) FORSYTH INST.
PA Smith DJ, Taubman MA;
PI
XX
DR WPI; 2003-845091/78.
XX
Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
XX
Claim 16; Page 15-16; 49pp; English.
XX
The present sequence is the protein sequence of Streptococcus downei glucosyltransferase-S (GTF-S). Peptide fragments of GTF-S, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Dieptopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus downei GbpB protein of the invention.
XX
SQ Sequence 1365 AA;

Query Match	100.0%;	Score 7230;	DB 9;	Length 1365;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1365;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MEKNRYKLHKVKKQWVAIGTITVLSFLAGGQVVAADTNNDGTSVQVNMKVPSPKPD	60	
DB	1	MEKNRYKLHKVKKQWVAIGTITVLSFLAGGQVVAADTNNDGTSVQVNMKVPSPKPD	60	
QY	61	AOAONGOLAQAMFKAANQADQATQTSQVSPATDGRVDNQVTPPAANQPAANQVANPAT	120	
DB	61	AOAONGOLAQAMFKAANQADQATQTSQVSPATDGRVDNQVTPPAANQPAANQVANPAT	120	
QY	121	DAGALNRQSAADTSTDGKAVPQTSQDPGHELETVYDANGQRLKNSYMWIDGKTTY	180	
DB	121	DAGALNRQSAADTSTDGKAVPQTSQDPGHELETVYDANGQRLKNSYMWIDGKTTY	180	
QY	181	FDGQTGEAQTDLPKTGOAQNDVPDSYQANNQAYSNEASSFETVDNYLTADSWYRPRKIL	240	
DB	181	FDGQTGEAQTDLPKTGOAQNDVPDSYQANNQAYSNEASSFETVDNYLTADSWYRPRKIL	240	
QY	241	KNGSQWQASSEGDLRPIILMTWPDAAATKAAYANFWAKEGLISGYRQNSANLDAATQNIQ	300	
DB	241	KNGSQWQASSEGDLRPIILMTWPDAAATKAAYANFWAKEGLISGYRQNSANLDAATQNIQ	300	
QY	301	SAIEKKIASEGTNWLRDKMSQVFKSQNQSIASENETVYPNODHMQGALLFSPNSKQTE	360	
DB	301	SAIEKKIASEGTNWLRDKMSQVFKSQNQSIASENETVYPNODHMQGALLFSPNSKQTE	360	
QY	361	HANSWRLLNRNPTFOTGKQKPYTTNYAGYELLANDVNSNPVQAEQLNHLHYLMWNG	420	
DB	361	HANSWRLLNRNPTFOTGKQKPYTTNYAGYELLANDVNSNPVQAEQLNHLHYLMWNG	420	
QY	421	DIVMGDKANFDGVRVDVAVDNNVADLLQIQDYKAKYGTQDQEKNAIDHLSILEAWSN	480	
DB	421	DIVMGDKANFDGVRVDVAVDNNVADLLQIQDYKAKYGTQDQEKNAIDHLSILEAWSN	480	
QY	481	DNDYVKDQNNFSLISINDQRSGMLKAFGVASAYRGNLSNATAGLKRSANPDSDPVNY	540	
DB	481	DNDYVKDQNNFSLISINDQRSGMLKAFGVASAYRGNLSNATAGLKRSANPDSDPVNY	540	
QY	541	VFIRAHDSVQTRIAKIIREKLGKTNADGLTNLTLDLANKAFDIYNQDMNATDKVYYPNN	600	
DB	541	VFIRAHDSVQTRIAKIIREKLGKTNADGLTNLTLDLANKAFDIYNQDMNATDKVYYPNN	600	
QY	601	LPMAVAMMLQNKDTVTRVYVYGDYTDNGQYMATKTPFYNAIETLLKGRIKYVAGGQAVSY	660	
DB	601	LPMAVAMMLQNKDTVTRVYVYGDYTDNGQYMATKTPFYNAIETLLKGRIKYVAGGQAVSY	660	
QY	661	KQDWSGILTSVRYGKGANSASDAGTETRNQSMALLINNRNFRAYRNLTLNMGAAHKS	720	
DB	661	KQDWSGILTSVRYGKGANSASDAGTETRNQSMALLINNRNFRAYRNLTLNMGAAHKS	720	
QY	721	QAYRPLLLSTKQGIATYLNDSVDVSRQYKTDQGNLSFASQVANAQVSGMTQWVY	780	
DB	721	QAYRPLLLSTKQGIATYLNDSVDVSRQYKTDQGNLSFASQVANAQVSGMTQWVY	780	
QY	781	PVGAADNQDVRTSPSTQATKQDNIHQSDALDSQVITYEGFSNFQAFQSPDQYTNVIAK	840	
DB	781	PVGAADNQDVRTSPSTQATKQDNIHQSDALDSQVITYEGFSNFQAFQSPDQYTNVIAK	840	
QY	841	NGDLFKSWGITQFEMAPQVVSSEDTGTFLLSVILNGYAFSDRYDLAMSKNKYGSQDLAN	900	
DB	841	NGDLFKSWGITQFEMAPQVVSSEDTGTFLLSVILNGYAFSDRYDLAMSKNKYGSQDLAN	900	
QY	901	AIKGLQSAQIKVLSLVLPNQLNLPQKEVVTATRVNQYQAKSGATINKTPVANTRSYG	960	
DB	901	AIKGLQSAQIKVLSLVLPNQLNLPQKEVVTATRVNQYQAKSGATINKTPVANTRSYG	960	

QY	961	DYQEQYGGKFLDDLOKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV	1020	
DB	961	DYQEQYGGKFLDDLOKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV	1020	
QY	1021	LSEGNKYLNLADGKFLPTVLNNTYQPOVSANGFTSKNGGIIHYLDKNGQEVKNRFEIS	1080	
DB	1021	LSEGNKYLNLADGKFLPTVLNNTYQPOVSANGFTSKNGGIIHYLDKNGQEVKNRFEIS	1080	
QY	1081	GSWYYFSDSGKMATGKTKIGNDTYLFMPNGKQJKEGVWYDGGKAYYDDNNGRTWTNKG	1140	
DB	1081	GSWYYFSDSGKMATGKTKIGNDTYLFMPNGKQJKEGVWYDGGKAYYDDNNGRTWTNKG	1140	
QY	1141	EFVRDQDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTVTINGKSYTFDADQGD	1200	
DB	1141	EFVRDQDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTVTINGKSYTFDADQGD	1200	
QY	1201	LVQTDNANPAPOGQAGWKLLGDNQWGYRKGQGLLTGEQITIDGQKVFQDNGVQVKG	1260	
DB	1201	LVQTDNANPAPOGQAGWKLLGDNQWGYRKGQGLLTGEQITIDGQKVFQDNGVQVKG	1260	
QY	1261	DASGVLRFYDRDQGHQVKGWYSTDNDNMYVNESQVLTGLQITIDGQTVYFDDKGIQAK	1320	
DB	1261	DASGVLRFYDRDQGHQVKGWYSTDNDNMYVNESQVLTGLQITIDGQTVYFDDKGIQAK	1320	
QY	1321	GKAVMDENGLRYFDADSGNMLRDRWKNVYDGNWYFNRNGLATRW	1365	
DB	1321	GKAVMDENGLRYFDADSGNMLRDRWKNVYDGNWYFNRNGLATRW	1365	
RESULT 3				
ID	AAU98029	standard; protein; 1430 AA.		
AC	AAU98029;			
XX				
DT	27-AUG-2002 (first entry)			
XX				
DE	S. mutans glucosyltransferase GTFD.			
XX				
KW	Glucosyltransferase; GTFD; transgenic plant; paper sizing;			
KW	coating composition; glucan; starch; latex; thermoplastic molecule;			
XX	amyloplast; vacuole; paper manufacture.			
OS	Streptococcus mutans.			
XX				
PN	US2002031826-A1.			
XX				
PD	14-MAR-2002.			
XX				
PF	19-DEC-2000; 2000US-00740274.			
XX				
PR	07-JUN-1995; 95US-00478704.			
PR	07-JUN-1995; 95US-00482711.			
PR	07-JUN-1995; 95US-00485243.			
PR	16-JAN-1998; 98US-00007999.			
PR	16-JAN-1998; 98US-00008172.			
PR	20-JAN-1998; 98US-00009620.			
PR	11-DEC-1998; 98US-00210361.			
XX				
PA	(NICH/) NICHOLS S E.			
XX				
PI	Nichols SE;			
XX				
DR	WPI; 2002-414332/44.			
DR	N-PSDB; ABK52940.			
XX				
PT	Glucosyltransferase B or D protein useful for producing a glucan useful			
PT	as substitutes for and additions to modified starch and latexes in paper			
PT	manufacture, comprises mutations in specific positions.			
XX				
PS	Disclosure; Page 38-42; 44pp; English.			
XX				
CC	The invention an isolated protein comprising a glucosyltransferase (GTF)			

DE	Streptococcus mutans glucosyltransferase-D.
XX	
KW	Glucosyltransferase; enzyme; vaccine; anticonaries; epitope; immunogen.
XX	
OS	Streptococcus mutans.
XX	
PN	W02003075845-A2.
XX	
PD	18-SEP-2003.
XX	
PF	07-MAR-2003; 2003WO-US006962.
XX	
PR	07-MAR-2002; 2002US-0363209P.
XX	
PP	08-AUG-2002; 2002US-0402483P.
XX	
PA	(FORS-) FORSYTH INST.
XX	
PI	Smith DJ, Taubman MA;
XX	
DR	WPI; 2003-845091/78.
XX	
PT	Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex Class II protein.
XX	
PS	Claim 16; Page 13-14; 49pp; English.
XX	
CC	The present sequence is the protein sequence of Streptococcus mutans glucosyltransferase-D (GTF-D). Peptide fragments of GTF-D, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein- binding peptide from S. mutants glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.
XX	
SQ	Sequence 1430 AA;
	Query Match 43.9%; Score 3174; DB 7; Length 1430; Best Local Similarity 47.3%; Pred. No. 5.3e-180; Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38
QY	1 MEKNRYKLHKVKKOWVAIGVT--VTLSFLAGGVVAADTNNDGTSV-----QVNKMV 53
DB	1 METKGRYKMKHKGHWTVAVASGLITLTTLTGSSVSAETEQTSDKVYTQSSEDDKAA 60
QY	54 PSDPKFDA---QAQNGLOAQMFKAAQADOTATSQVS PATDGRVDNVT PAANQPAAN 109
DB	61 SSSQTDAPKTQATEQ-TQAQ-SQANVAD-TSTG----ITKETPSQNITTQANSDDKT 113
QY	110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDFG-----H 149
DB	114 VNTKSEEAQTEERTKQSEEAQTASSQALTQAKAE LTKRQTAAENKNPVDLAAIPN 173
QY	150 LETVDGKYTYVDANGORLKYNYSWIDGTYTFDQGTGEAQTDLP-----KTGANQDNVPD 205
DB	174 VKQIDGKYTYTGSDDQPKNFALTVNNKLVYFDKNTG-ALTDTSQYQFKOGLTKLAN--- 229
QY	206 SYQANNOAYSNEASSFEVDNYLTADSNYRPRLIKNGOSQWASSEGDLPILMTWPDA 265
DB	230 DTPHNQIVNFTSLETIDINVTADSNYRPDLKNGKWTASSESDLPLMLSWDPK 289
QY	266 ATKAAAYANFWAKEBLISG-SYRONSA--NLDAATQNIQSAIEKKIASEGNTNWLDRKMSQ 322
DB	290 QTOIAYLNVMNQGLGTGENYTDSSQESLNLAQTQVVKIETKISQTOQTQWLRIINS 349
QY	323 FVKSQNWQSIASENETVYPNDHQGGALLFSNSKOTEHANSDRWLLRNRPFTQCKQKY 382
DB	350 FYKTOPNWSQTESAGEKHLEQGALLYSNSDKTAYANSDYRLINRPTTSQTKPKY 409

XX AC ADX37279;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus mutant glucan binding protein B variant #8.

XX KW immunogenicity; immune stimulation; glucan binding protein-B;

XX KM microparticle; major histocompatibility complex; tooth disease.

XX OS Streptococcus mutans.

XX PN US2005031633-A1.

XX PD 10-FEB-2005.

XX PF 09-MAR-2004; 2004US-00797821.

XX PR 13-APR-1998; 98US-0081550P.

XX PR 08-JAN-1999; 99US-0115142P.

XX PR 12-APR-1999; 99US-00290049.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PR 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.

XX (TAUB/) TAUBMAN M A.

XX Smith DJ, Taubman MA;

XX WPI; 2005-151644/16.

XX New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.

XX Claim 7; SEQ ID NO 36; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a Streptococcus mutans GbpB protein of the invention.

XX SQ Sequence 1430 AA;

Query Match 43.9%; Score 3174; DB 9; Length 1430;

Best Local Similarity 47.3%; Pred. No. 5.3e-180;

Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

QY 1 MEKNRYKLHKVKQWVAIGVTT--VTLSPLAGQGVVAADTNNDGTSV-----QVKNKV 53

DB 1 METKRYKHKVKWVAVASGLITLTGTLGSSVSAETEOQSDKVVTKSEDDKAA 60

QY 54 PSDPKFDA----QAQNGQLAQMFKAAQADQATQTSQVSPATDGRVDNVTPAANPAAN 109

DB 61 SESSQTDAPKTKQAQTEQ-TQAQ-SQANVAD-TSTS----ITKETPSQNIITTOANSDDKT 113

QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDPG-----H 149

DB 114 VTMKSEEAQTSERTKQSEEAQTATSSQALTOAKAELTKQRTAAQENKNPVDLAAIN 173

QY 150 LETVDGKTYVDANGORLKNYSVIDGKTYVPDQTGEAQTDLP-----KTGQANQDNVPD 205

DB 174 VRQIDGKYIYIGSDGPKKKNFALTANKVLYFDKNTG-ALTDTSQYQPKQGLTKLNN--- 229

QY 206 SYQANNOAVSNEASSPETVDNYLTADSWTRPKILKNGSQSWOASSGDLRPLMTWPD 265

DB 230 DYPHNIQVNFENTSLETIDNYVTADSWTRPKDILKNGKVTATSSDLRPLMSWPDK 289

QY 266 ATKAAVANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIECKIASSEGNTNWLRDKMSQ 322

DB 290 QTOIAYLNMNQOGLGTGENTADSSQESLNLAAQTQVKIETKISQTOQTQWLADINS 349

QY 323 FVKSQNSIASENETVYPNOHMQGALLFNSKDTBHANSDWRLNRPFTQTGKQKY 382

DB 350 FVKTPNWNSTESDTSAGEKHLQCGALLYSNSDKTAYANSDYRLLNRPFTSQTGPKY 409

QY 383 FTTNYA-GYELLANDVNSNPVQAEOLNHLHYLMNMGDI VMGDKDANFDGVRDVAVDN 441

DB 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNWLHYLMNYSIVANDPEANFDGVRDVAVDN 469

QY 442 VNADLLQIQORDYYKAKYGTQONEKNAIDHLSILEAWSGNNDNDYVKDQNNFSLSDNDORS 501

DB 470 VNADLLQIATSDYLKAHYGVDKSEKNAIHLISILEAWSNDNDPOYKDKTKGAQLPIDNKLRL 529

QY 502 GMLKAF-----GYASAYRGNLSNLATAGLNKNSA-NPDSDPVNPVYFIRAHSEVQT 552

DB 530 SLLYALTRPLEKDNASKNEIRSGLEPVITNSLNNRSEAGKNSERMANVIFIRAHSEVQT 589

QY 553 RIAKIIREKLKGTNADGLTNLTDLNKAFTDYNODMNATDKVYYPNNLPMAAYAMLQNK 612

DB 590 VIAKIIKAQI-NPKTDGLT-FTLDELKQAFKLIYNEDMRQAKKCYTQSNIPYAYALMWSNK 647

QY 613 DTVTRVYYGDMYTDNGQYMATKTPFYNAIETLLKGRIKYVAGCQ--AVSYKQ-----DW 664

DB 648 DSITRLYYGDMYSDDGQYMATKSPYDAIDTLKKARIKYAAGQDMKITYVEGDKSHMDW 707

QY 665 S-SGILTSVRYGKGANSADAGTETRNMGALLNNRPNFRAYRN--LTLNMGAAHSQ 721

DB 708 DYTGLVTSVRYGTGANETDQGEATKTQGMVITSNPNLSKLNQNDKVINVMGAHKNQ 767

QY 722 AYRPLLSTKGIATYLNDSVDVSQYKYTDSQGNLSFASSELSQVANAQVSGMTQVWPV 781

DB 768 EYRPLLTTKQGLTSYSDAAKS-LYRKTNDKGELVDFDASDIQGLYLPQVSGYLAVWP 826

QY 782 VGAAQNDQVRSPTSQATKQGNIHOSDALDSQVIYEGFSNFQAFQSPDOQTVNAVIKN 841

DB 827 VGASQNDQVRAASKANATQGVYESSALSQLIYEGFSNFQDFVTKDSDYTNKKIAQN 886

QY 842 GDLFSWGITOPEMAPQVSSSDGTFLDSVLNGYAFSDRYDLAMSKNKYSGKODLANA 901

DB 887 VQLFWSWGTSPFEMAPQVSSSDGSLDSIIQNGYAFEDRYDLAMSKNKYSGQODMINA 946

QY 902 IKGLQSAQIKVLSDLPVNLQNLPEKEVVTATRVNQYQAKSGATINKTPYVANTRSYG- 960

DB 947 VKALHKSQIQVIADWVPQIYNLPKEVVTATRVNDYGEYKDSSEIKNTLYAANTKSNQK 1006

QY 961 DYQEOYGGKFLDDLQKLYPRLPSTKQISTGKPIDPSVKITNWSAKYFNINGSNILGRGAKYV 1020

DB 1007 DYQAKYGAFSLSELAAKYPSIPNRTQISNGKKIDSEKLTAMKAKYFNGTNILGRGVYV 1066

QY 1021 LSE--GNKYLNIADGKLFPTVLNNYQOQVPSANGFISKNGGIHYLDKNGQEVKNRF-K 1077

DB 1067 LKDNASDKYFELKGNQYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFYQ 1121

QY 1078 EISGSWYFYFSDCKMATGKTKIGNDTYLFMPNGKOLKEGVW--YDGKAYYYDDNGRTWT 1135

DB 1122 DAGKNWYFYDNNGHVYGLQNGEVQYFLSNGVQLRESFLENADGSK-NYFGLGNRYS 1180

QY 1136 NKGFEFRVGDQKVRYPNGDGTIAIGLSVLNDRLYFDAYGYQVKGQTVT-INGKSYTF 1194

DB 1181 N-GYYSF--DNDSKRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITSDGKRYF 1237

QY 1195 DADQGLVQTDNVANPAQOAGKLLGDNQWGY-RKDQGLLTGEQITDQKQVFPDNGVQ 1253

DB 1238 DDGSGNMAVNRFANDK-----NGDWYILNSDGIALVGQVQTINGKTYFYFGDQGX 1286

QY 1254 VKGQTATDASGVLYRFDQHQVKGWYTSDDNNWYVNESQVLTGLQITDGTQVTFD 1313

DB 1287 IKGKIITD-NGKLKYFLANSGLARNIFATDSQNNWYFGSDGVAVTQSTTAGKKLYFA 1345

QY 1314 DKGIQAKGKAVMDENGNLRYFDADSGNMLRDRWK-NVDGNWYFYNNRNGLA 1362

Db 1346 SDGQVKGSPV-TYNGKRVHYHADSGELQVRFEADKGNWYYLDNGEA 1394

RESULT 6

AAU98043

ID AAU98043 standard; protein; 1430 AA.

AC AAU98043;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFD mutant N471D.

XX Glucosyltransferase; GTFD; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 471 /note= "Wild-type Asn substituted by Asp"

FT US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE,

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from 1448V, D457N, D567T,

CC K104T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T, D567T/D571K/K1014T,

CC I448V/D457N/D567T/D571K/K799Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes B or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked to a promoter, a vector comprising the expression cassette, host

CC cell introduced with the vector, a transgenic plant comprising the

CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or

CC coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GTF, wild type or, starch, a latex,

CC thermoplastic molecule or their combinations or glucan and starch where

CC the glucan is produced in the amyloplast and/or vacuole or a maize line

CC deficient in starch biosynthesis, transformed with a gene encoding a

CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

CC comprising the glucan (paper sizing/coating agent). The vector is useful

CC for producing a glucan in a plant. The method comprises transforming a

CC plant cell with the vector, growing the plant cell under plant growing

CC conditions to produce a regenerated plant and inducing expression of the

CC polynucleotide for a time sufficient to produce the glucan in the

CC regenerated plant, where the vector contains a transit sequence from

CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and

CC chlorophyll AB binding protein to produce a transgenic plant, and glucan

CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC Glucans are useful as substitutes for and additions to modified starch

CC and latexes in paper manufacture. Unlike prior art techniques, which

CC require input materials that produce chemical effluents, paper

CC manufacture utilising the glucan produced by GTF, which utilises

CC biologically produced input materials, is more cost-effective and

CC environmentally friendly. Moreover, glucans also exhibit thermoplastic

CC properties and impart gloss to the paper during coating step. The present

CC sequence represents a GTFD mutant of the invention. Note: The present

CC sequence is not shown in the specification but was created by the indexer

CC using the GTFD sequence appearing as AAU98029 and the information in

CC claim 36

XX Sequence 1430 AA;

SQ

Query Match 43.8%; Score 3169; DB 5; Length 1430;

Best Local Similarity 47.3%; Pred. No. 1.1e-179;

Matches 676; Conservative 214; Mismatches 436; Indels 104; Gaps 38;

QY 1 MEKNRYKHLKVKQWVAIGVTT--VTLSFLAGGVAAADTNNNDGTSV-----QVNMV 53

DB 1 METKRYKMHKVKGHVTVAVASGLITLGTTLTGSSVSAETEQTSDKVVTKSDDKAA 60

QY 54 PSDPKFDA---QAQNGOLAQAMFKAANOADOTATSOVSPATDGRVDNQVTPAANOAPAAN 109

DB 61 SESSQTDAPKTKQAQTEQ-TQAQ-SOANVAD-TSTS-----ITKETPSQNIITQANSDDKT 113

QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDQPG-----H 149

DB 114 VINTKSEEAQTESEETKQSEEAQTASSQAQTQAQAEKLTQKQTAQENKNPVOLAIPN 173

QY 150 LETVGGKTYVDANGORLKNYSWVIDGKTYYPFDGTEAQTDLTP-----KTGAQNDVNDP 205

DB 174 VKQIDGKTYIIGSDGQPKKNFALTVNNKVLVFDKNTG-ALTDTSQVQPKGLTKLNN--- 229

QY 206 SYQANNOAYSNEASSFETVDNVLTDADSWYRPRKILKNGSQWASSEGDLRPLMTWPPA 265

DB 230 DYTPHNQIVNFENTSLKETIDNVYTDADSWYRPRKILKNGKTKTASSEDRLPLMSWPPDK 289

QY 266 ATKAAYANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEKKIASEGNTNWLDRKMSQ 322

DB 290 QTQIAYLVNMQOGLGTGENTADSSQSLNLAQTQVQVKIETKISQTOQTQWLRIINS 349

QY 323 FVKSQNSQNSIASENETVYPNQDHMQGALLFNSKDTSHANSWRLNLRNPTFTQTKQKY 382

DB 350 FVKTPNNNSQTESDTSAGEKDLQGGALLYSNSDKTAYANSYRLLNRTPTSTQTKPKY 409

QY 383 FTTNYA-GYELLANDVNSNPVQAEQINHLHYLMNGDIIVMGDKDANFDGVRVDVND 441

DB 410 FEDNSSGGYDFLLANDIDNSNPVQAEQINHLHYLMNYGSIIVANDPEANFDGVRVDVND 469

QY 442 VNADLLQIQDYRYKAKYGTQDNEKNAIDHLSLTLEAWSGNDNDYVKDQNNFSLINDORS 501

DB 470 VDADLLQIASDYLKAHYGVDSKNAINHLISLEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529

QY 502 GMLKAF-----GYASAYRGNLSNLATAGLKNRSA-NPDSDPVNTVYFIRAHSEVQT 552

DB 530 SLLVALTRPLEKDSANKNKNEIRSGLEPVITNSLNRSAEGKNSERMANVIFIRAHSEVQT 589

QY 553 RIAKIIRKLGKTNADGLTNTLDDLKAFDIYNQDMNATDKVYPPNKLPMAYAWMLQNK 612

DB 590 VIAKIIKAQI-NPKTQDGLT-FTLDELKQAFKIYNEDMRQAKKKTQSNIPYALMLSNK 647

QY 613 DTVTRVYVGMVDTNGQVMATKTPPYNAIETLLKGRIVYAGGO--AVSYKQ-----DW 664

DB 648 DSIRKLYVGMVSDGQPMATKSPYDAIDITLLKARIKYAAGGQDMKITYVEGDKSHMDW 707

QY 665 S-SGLTISVRYKGGANSADAGNTETRNSGMLLINRPNFRAYRN--LTLNMGAAHKSQ 721

DB 708 DYTGLTISVRYGTGANEATDQSGSEATKTQGMVITSNNSPLKLNQNDKVINVMGAAHKNQ 767

QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDPQG-----H 149
DB 114 VNTKSEEAQTSERTKQSEEAQTASSQALTOAKAELTKQRTQAAQENKNPVDLAAIEN 173
QY 150 LETVDGKTYVDANGORLKNYSWVIDGKTYVFDGQTGEAQTDLF-----KTGQANQDNVPD 205
DB 174 VKQIDGKYIYGSDGQPKNFALTUVNKNVLYFDKNTG-ALTDTSQYQFQGLTKLNN--- 229
QY 206 SYQANNOAYSNEASSFETVDNYLTADSWYRPRKILKNGSQWQASSECDLRPIILMTWPPDA 265
DB 230 DYTPHNQIVNFENTSLETIDNYVTADSWYRPRKILKNGKTWTWTASSSDLRPLMSWWPDK 289
QY 266 ATKAAAYANFAWEKLLISG-SYRONSA--NULDAATONIQAIEKKIASSEGNTNWLKMSQ 322
DB 290 QTQIAYLNTMNOQGLGTGENYATDSSQESLNLAAQTQVVKIETKISQTQOTQWLRLDIINS 349
QY 323 FVKSQNOWSIASNETVYPNQDHMOGALLFNSKDTEHANSWRLLNRPFTQCKQKY 382
DB 350 FVKTPNWSQTESDTSAGEKHLOGGALLYSNDSKTAYANSDYRLNRTPTSTQCKPKY 409
QY 383 FTTNYA-GYELLILANDVDNSNPVQAEQLNHLHLYLNNWGDIVMGDKDANFDGVRVDAVDN 441
DB 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNHLHLYLNNYSIVANDPEANFDGVRVDAVDN 469
QY 442 VNADLLQIORDYYKAKYQTDQNEKNAIDHLSILEAWSGNDNDYVKDQNNFSISIDNDQRS 501
DB 470 VNADLLQIASDYLIKAHYGVDKSEKNAINHLSILEAWSNDNDPQVNDKTKGAQLPIDNKLRL 529
QY 502 GMLKAF-----GYASAYRGNLSNLATAGLKNRSA-NPDSDPVPNYVFIKRAHDSVQT 552
DB 530 SLTYALTREPLEKADSNKEIRSGLEPVITNSLNRSAEGKNSERMANYIFIRAHDSVQE 589
QY 553 RIAKIIREKLGKTNADGLTNLTDLLNKAFDIYNQDMNATDKVYPNNLPMAWMLQNK 612
DB 590 VTAKIIKAQI-NPKTDGLT-FILDELKQAFKIYNEDMRQAKKYTQSNIPYALMLSNK 647
QY 613 DTVTVYYGDMYTDNGQYNAKTPFYNAIETLLKGRIKYVAGQ--AVSYKQ-----DW 664
DB 648 DSITRLYYGDMYSDGQYNAKTPFYDAIDTLKARIKAYAGQDMKITYVEGDKSHMDW 707
QY 665 S-SGILTSVRYCKGANSASDAGNTEFRNSGMALLNNRNFPRAYRN--ITLNGAAHKKQ 721
DB 708 DYTGLVTSVRYGTGANEATDQSEATKTQGMVITSNPSSLKLNQNDKVINVNGAAHKNQ 767
QY 722 AYRPLLSTKQGIATYLNDSVDROKYKVTDSQGNLSFASSELQSVANAQVSGMTQVWPV 781
DB 768 EYRPLLTTKQGLTSTYSDAAKS-LYRKTNDKGLVFDASDIQGLNPOVSGYLAWVP 826
QY 782 VQAADNQDVRTSPSTQATKQGNIIHQSDALDSQVIYEGFSNFQAFQAQSPDQYTNVIAKN 841
DB 827 VQASNDQVRVAASNKANATQVYESSSALDSQLIYEGFSNFQDFVTKDSDYTNKKIAQN 886
QY 842 GDLFKSWGITQEMAPQVYSSDGTFLDSVIINGVAFPSDRYDLAMSKNKYGSKODLANA 901
DB 887 VQLFKSWGITSFEMAPQVYSSDGTFLDSIIQNGVAFEDRYDLAMSKNKYGSKQDMINA 946
QY 902 IKGLQAGIKVLSDLPVNLNPKGEVVTATRVNQYGOAKSGATINKTPYVANTRSYG- 960
DB 947 VVALHKSGLQVADWVFDQIYNLPKGEVVTATRVNDYGEYRDKSEIKNTLYAANTKSNK 1006
QY 961 DYQEQYGGKFLDLOKLQYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKTV 1020
DB 1007 DYQAKYGGAFLESLAAKYPISFNRTQISNGKKIDFSEKITAMKAKYFNGTNILGRGVGV 1066
QY 1021 LSE--GNKYLNLADGKLFPTVLNNTYQOPQVSANGFISKNGGHIHVDKNGQEVKNRP-K 1077
DB 1067 LKDNASDKYFELKGNQTLPLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
QY 1078 ELSGSMWYFDSGKWATKTKIGNDTYLPMPNGKOLKEGW--YDGKAYYYDDNGRTWT 1135
DB 1122 DAGKNWYIFDNNHVMYGLQNGEVQYPLSLNGVQLRESFLENADGSK-NYFCHLGNRSY 1180
QY 1136 NKGVEFRVDGQDKWRYFNGDGTIAIGLVSLDNRLTYFDAYGYQVKQGTVT-INGKSYTF 1194

DB 1181 N-GYTSF--DNDSKWRFYDASGVMAVGLKTINGTQYFDQDGYQVKGAMITGSDGKRYF 1237
QY 1195 DADQDGLVQTDNANPAPOGAGKLLGDNQWGY-RKDQGLLTGQEQITDGQKVFQDNGVQ 1253
DB 1238 DGSNGMAVNRFPANDK-----NGDWIYLSNDSGIALVGVQTINGKTYTFGQDQKQ 1286
QY 1254 VRKGATDASGVLRFYDRDQGHQVGKMYSTSDNNVYVYNSQVLTGLQITIDGQTVYFD 1313
DB 1287 IKGKIITD-NGKLKYFLANSAGELARNIFATDSQNNWYFSGDSGVAVTGSQTIAGKLYPA 1345
QY 1314 DKGIOAKGVAVDENGNLRYFADSGNMLRDRWK-NVDGNWYFENRGLA 1362
DB 1346 SDGQVKYGSFV-TYNGKVHYHADSGELQVNRFPADKGNWYIYLDNGEA 1394
RESULT 8
AAU98041
ID AAU98041 standard; protein; 1430 AA.
XX
AC AAU98041;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFD mutant T589D.
XX
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
XX
EH Key Location/Qualifiers
FT Misc-difference 589 /note= "wild-type Thr substituted by Asp"
XX
XX US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
PA (NICH/) NICHOLS S E.
XX
FI Nichols SE;
XX
DR WPI; 2002-414332/44.
XX
PT Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
PS Claim 36; Page; 44pp; English.
XX
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host

cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GFP, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GFP, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GFPD mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GFPD sequence appearing as AAU98029 and the information in claim 36

XX
SQ Sequence 1430 AA;

Query Match 43.8%; Score 3168; DB 5; Length 1430;
Best Local Similarity 47.3%; Pred. NO. 1.2e-179;
Matches 676; Conservative 213; Mismatches 437; Indels 104; Gaps 38;

QY 1 MEKNRYKLHKVKKVAVLGVTT--VTLFLAGGVAAADTNNDGTSV-----GVNKV 53
DB 1 METKRYKHKVKKVHWTVAVASGLITLGTTLGSSVSAETEQDTSKVVTKSDDKAA 60
QY 54 PSDPKFDA---QAQNGLOAAMFKAANOADOTATSOVSPATDGRVDNQTVAANOQPAAN 109
DB 61 SESSQTDAPKTKQAQTEQ--TQAQ--SQANVAD--TSTS---ITKETPSQITTAQNSDDKT 113
QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQPG-----H 149
DB 114 VNTKSEEAQTSBERTKQSEEAQTASSQALTKQAELTKQRQTAQENKNPVDLAAIN 173
QY 150 LETVDGKTYVDANGQRLKNYSVIDGKTYVPDGTGEAQTDLP-----KTQANQDNVPD 205
DB 174 VKQIDKRYIIGSDGQPKNFALTNNKVLVFDKNTG-ALTDTSQYQFKGLTKLNN--- 229
QY 206 SYOANNOAYSNEASSPETYVDNYLTADSWYRPRKILKNGOSQWASSGDLRPLMTWPDPA 265
DB 230 DYTPHNQIVNFENTSLETIDNYVTADSWYRPRKILKNGKTWTASSSDRLPLMSWPDK 289
QY 266 ATKAAVANWAKEGLISG--SYRONSA--NLDATQNIQSAIEKKIASBGTWHLRDKMSQ 322
DB 290 QTQIAYLNTYMNQOGLGTGENYTRADSSQESLNTAAQTQVVKIETKISQTOQTOWLRDIINS 349
QY 323 FVKSQNOWSIASNETVYPNQDHMOGQALLFNSN KDTEHANSQWRLNLRNPFTQTKQKY 382
DB 350 FVKTPQWNSQTESDTSASEKHLOQGALLYNSDKTAYANSQYRLLNRTPTTSQTKPKY 409
QY 383 FTTNYA-GYELLANDVDNSNPVQAEQLNHLHYLMNWGDI VNGDKDANPDGVRVDAVDN 441
DB 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNHLHYLMNWGDI VANDPEANPDGVRVDAVDN 469
QY 442 VNADLLOIQRDYKAKYCTDQEKNAIDHLSLEAWSNDNDYVKDQNNFSLIDNDQRS 501
DB 470 VNADLLOIASDYLKAIHYGVDKSEKNAIHLISLEAWSNDNDPQNKDTKGAQLPIDNKLRL 529
QY 502 GMLKAF-----GYASAYRGNLSNLATAGLKNRSA--NPDSPPVNPYVFIHAHSEVQT 552

DB 530 SLLYALTRPLEKASKNKNEIRSGLEPVTITNSLNNRSAEGKNSERMANYIFIRAHSEVQD 589
QY 553 RIAKIIRKELKTNADGLTNLTLDLNKAFDIYNQDMNATDKVYYPNNLPMAYAWLQNK 612
DB 590 VIATIKAKI-NPKTDGLT-FTLDELKQAFKIYNEDMRQAKKKTQSNIPPTAYALMSNK 647
QY 613 DVTREVYGDVMTDNGQYMATKTPYNAIETLLKGRIVYAGGQ--AVSYKO-----DW 664
DB 648 DSITRLYIGDMYSDDGQYMATKSPYDAIDTLKARIKYAAGGQDMKIITYVEGDSKMDW 707
QY 665 S-SGLTSVRYGKGANSASDAGNTRNSGSMALLINRPNFRAYRN--LTLNMGAAHSQ 721
DB 708 DITGVLTSVRYGTGANETDQSEATKTCGMAVITSNPSLKLNQNDKIVNMGAAHKQ 767
QY 722 AYRPLLSTKQIATYLNDSVDSDSOYKYKTDQSGNLSFASSELQSVANAQSGMTQVWVP 781
DB 768 EYRPLLSTKQIATYLNDSVDSDSOYKYKTDQSGNLSFASSELQSVANAQSGMTQVWVP 826
QY 782 VGAAADNQVRTSPSTQATKQGNIVHQSALDSOVIYEGFSNFQAFQAQSPDQVNAVIQ 841
DB 827 VGASDNQDVRVAASNAKANATQGVYESSALDSQLIYEGFSNFQDFVTKDSDTNNKIAQN 886
QY 842 GDLFKSWGITOPEMAPQVSSDGTFLDSVILNGYAFSDRYDLAMSKNKYGSKQDLANA 901
DB 887 VQLFKSWGITOPEMAPQVSSDGTFLDSVILNGYAFSDRYDLAMSKNKYGSKQDLANA 946
QY 902 IKGLOSAGIKVLSDLPVNLQNLNLPKEVVTATRVNQYQAKSGATINKTPPYANTRSYG- 960
DB 947 VKALHKSQIQTADVWVPQIYNLPKEVVTATRVNDYGEYRKDSBIKNTLYAANTKSNK 1006
QY 961 DYQEOYGGKFLDLQKLYPRLPSTKQISTGKPIDPSVKITNNSAKYFNGSNILGRGAKTV 1020
DB 1007 DYQAKYGAFLSELAAKYPSIFNRTQISNGKIDSEKITAWAKYFNGTNILGRGVTV 1066
QY 1021 LSE--GNKYLNLADGKFLPTVLNNTYQOPQVSAANGFISKNGGIIHYLDKNGQEVKNRF-K 1077
DB 1067 LKDNASDKYFELKGNQTYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
QY 1078 EISGSWYFSDSGKAWATKTKIGNDTYILFMPNGKQLKEGVW--YDGKAIYYDDNGRTWT 1135
DB 1122 DAKGNWYFDDNNGHVMYGLQNGEVQVPLSNGVQLRESFLENADGSK-NYFGLHGNYS 1180
QY 1136 NKGFEVERVDGQKWRVFNCGDTIAGLVSLNRLTYPDAYGVQVKGQTVT--INGKSYTF 1194
DB 1181 N-GYTSF--DNDSKRYFYDASGVMAVLKTINGTQYFDQDGYQVKGAWITSGDKKRYF 1237
QY 1195 DADQGDLYQTDNANPAQOAGWKLIGDNQWGY-RKDGQLLTGSEQTIDGQKVFQDNGVQ 1253
DB 1238 DDGSGNMAVNRFANDK-----NGDWYLLNSDGIALVGQTINGKTYVFGQDGKQ 1286
QY 1254 VKGGTATDASGVLYRFDYRDOGHQVKGWYSTDDNHWYVNESGOVLTLGLOTTDQGVTFD 1313
DB 1287 IKGKIITD-NGKLYFLANSGLARNIIFATDSQNNWYFPGSDGVAVTGSGTTAGKLYFA 1345
QY 1314 DKGIOAKGKAVDENGRLRYPDADSGNMLRDRWK-NVDGNWYFNRNGLA 1362
DB 1346 SDGKQVKGSVF--TYNGKVHYHADSGELQVNRFEADKQGNWYLLDSNGEA 1394

RESULT 9
AAU98044
ID AAU98044 standard; protein; 1430 AA.
XX
AC AAU98044;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GFPD mutant N471D/T589D.
XX
KW Glucosyltransferase; GFPD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX

OS Streptococcus mutans.
XX Synthetic.

Key Location/Qualifiers

FT Misc-difference 471 /note= "Wild-type Aan substituted by Asp"
FT Misc-difference 589 /note= "Wild-type Thr substituted by Asp"

PN US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

PR 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I48V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilises the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36

XX	SQ	Sequence 1430 AA;
		Query Match 43.7%; Score 3163; DB 5; Length 1430;
		Best Local Similarity 47.2%; Pred. No. 2.4e-179;
		Matches 675; Conservative 214; Mismatches 437; Indels 104; Gaps 38;
QY	1	MEKNRYKLGKHKVQWVAIGVTT--VTLSFAGGQVVAADTNNNDGTSV-----QVNMV 53
DB	1	METKRRYKHKVKKHWTVAVASGLITLTGTTTLLGSSVSAETEQTQSKVVTQKSEDDKAA 60
QY	54	PSDPKFDA-----QAQNGOLAQAMPKAAQADQATQTSQVSPATGRVDNQVTPPAANQPAAN 109
DB	61	SESSQTDAPKTKQAQTEQ-TQAQ-SQANVAD-TSTS-----ITKETPSQNTITQANSDDKT 113
QY	110	VANQDVANPATDAGALNROSAADTSTDGKAVP-----QTSQDPG-----H 149
DB	114	VNTKSEEAQTSEERTKQSEEAQTASSQALTKQAKAELTKQRTAAQENKKNPVDLAAIPN 173
QY	150	LETVDGKTYVVDANGQRLKNYSMVIDGKTYFDFGQTGEAQTDLF-----KTQANQDNVDP 205
DB	174	VKQIDGKYVYIGSDGQPKKNFALTAVNKVLYFDKNTG-ALTDTSQYQFKQGLTKLNN--- 229
QY	206	STQANNQAYSNEASSPFTVDNYLTADSWTRPRKILKNGSQWQASSEGDLPILMTWMPDA 265
DB	230	DYTPHNQIVNFENTSLETIDNYVTADSWTRPRKILKNGKWTWASSSDLRPLLSWWPDK 289
QY	266	ATKAAVANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEKKIASSEGNWLRDKMSQ 322
DB	290	QTQIATLNTNMOOGLTGENTYADSSQESLNLAAQTQVKIETKISQTOQTQWLRLDINS 349
QY	323	FVKSQNWSIASENETVYPNQDHMOGQALLFNSKDTSEHANSWRLNLRNPTFTQTKQKY 382
DB	350	FVKTPQWNWSQTESDTSAGEKHLQGGALLYNSDKTAYANSDYRLNRTPTTSQTGPKPY 409
QY	383	FTTNVA-GVELLIANDVNSNPVQABQLNHLHYLANMGDIVMGDKDANPDGVRVDAVDN 441
DB	410	FEDNSSGGYDFLLANDIDNSNPVQABQLNHLHYLANMGDIIVANDPEANPDGVRVDAVDN 469
QY	442	VNADLLOIORDYKAKYGTQDQEKNAIDHLSLTLEAWSGNDNDYVQDNFSLSDNDQRS 501
DB	470	VADLLQIATSDYLKHAHYGVKSEKNAINHLSLTLEAWSNDNDPQYNKDTGAQLPIDNKRL 529
QY	502	GMLKAF-----GYASAYRGNLSNLATAGLKNRSA-NPDSPPVNPYVYFIRAHDEVOT 552
DB	530	SLLYALTREPLEKXADSKNEIRSGLEPVTNSLNRSAEKGNSERMANYIFIRAHDEVQD 589
QY	553	RTAKIIREKLGKTNADGLTNLTLDDLKAFDYNQDMNATDKVYYPNNLPMAWMLQNK 612
DB	590	VIAKIITKAQI-NPKTDGLT-FTLDELKQAFKIYNEQDMRQAKKKTQSNIPYALVMSNK 647
QY	613	DTVTRVYGDMDYTDNGOYMATKTPFYNAIETLLKGRIKYVAGGQ--AVSYKQ-----DW 664
DB	648	DSITRLYGDMDYDSDGQYMATKSPYDAIDTLKARIKYAAGGQDMKITYVEGDKSHMDW 707
QY	665	S-SGILTSVRYGKANSASDAGNTETNSGMALLNNRNPFRAYRN--LTLNMGAAHKSQ 721
DB	708	DYTGVLTSVRYTGANEATDQSEATKTQGMAVITSNPSLKNQNDKVIWVMGAHKNQ 767
QY	722	AYRPLLLSTKGIATYLANDSDVDSDRQYKYQYTDQSGNLSFSASBELQSIVANAQVSGMIQVWP 781
DB	768	EYRPLLLATKDLTSYSDAAKS-LYRKTNDKGLVFDASDIQGYLNPQVSGYLAVWP 826
QY	782	VCAANDQVRTSPSTQATKQDGNHYHQSDALDSQVLYEGFSNFQAQSPDQTVNAIAGN 841
DB	827	VGASNDQVRAASNAKANATQGVYESSALDSQLIYEGFSNFQDFVTKDSDTNNKIAQN 886
QY	842	GLFKSWGITQEPMAPOVYSSSDGTFLDSVILNGVAFSDRYDLAMSKNNKYSKODLANA 901
DB	887	VOLFSGWGTSEMAPQVYSSSDGSLFDSIIQNGVAFEDRYDLAMSKNNKYSQQQMINA 946
QY	902	IKGLQASAGIKVLSLDLPVNLPGKEVVVTAATRVNQYQAKSGATINKTPPYVANTRSYG- 960

Db 947 VKALHSGIOVIADWVPDQIYNLPCKEVVATRVNDYGEYKDSBKNTLYAANTKSNCK 1006
QY 961 DYQEQYGGKFLDLDQKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 1007 DYQAKYGGAFLELAAKYPSIFNRTQISNGKKIDPSEKITAMKAKYFNGTNILGRGVGV 1066
QY 1021 LSE--GNKYLNLADGKLFPLTVLNTYQOPVSANGFISNGGIHYLDKNGQEVKRP-K 1077
Db 1067 LKDNASDKYFELKGNQTYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFPVQ 1121
QY 1078 EISGSMYYFSDSGKMATGKTKLGNDRYLFMPNGKQLKEGVW--YDGKKAYYYDDNGRWTW 1135
Db 1122 DAKGNWYFDNNGHMYVGLQNGEVOYFLSNGVQLRESFLENADGSK-NYFCHLGNRY 1180
QY 1136 NKGFEVFRVGDGKRWYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTVT-INGKSYTF 1194
Db 1181 N-GYYSF--DNDSKRWYFDSAGVMAVGLKTNGNTQYFDQGVQVKGAWITSGDKGRYF 1237
QY 1195 DADQGLVOTDNANPAQOAGCKWLLGDNQWGY-RKDQGLLTGEGTIDQKVFQDNGVQ 1253
Db 1238 DDGSGNMAVNFANDK-----NGDWYLLNSDGIALVGQVTNGKTYFFGQDGKQ 1286
QY 1254 VKGGTATDASGLVRFVDRDQGHQVGKGYSTSDDNWYVNESQVLTGTQIDGOTVYFD 1313
Db 1287 IKGKIITD-NGKLKYFLANSGLARNIFAIDSONWYFPGSDGVAVTGSQTTIAGKRLYFA 1345
QY 1314 DRGIAQKAKAVMDENGLRYFDADSGNMLDRWK-NVDGNWYFNRNGLA 1362
Db 1346 SDGKQVKGFSV-TYNGKVHYHADSGELQVNRFEADKQGNWYLLDSNGEA 1394

RESULT 10

ID AAU98045 standard; protein; 1430 AA.

XX AAU98045;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFD mutant N471D/T589E.

XX Glucosyltransferase; GTFD; transgenic plant; paper sizing;

XX coating composition; glucan; starch; latex; thermoplastic molecule;

XX amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

XX Synthetic.

XX Key Location/Qualifiers

XX FT Misc-difference 471 /note= "Wild-type Asn substituted by Asp"

XX FT Misc-difference 589 /note= "Wild-type Thr substituted by Glu"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

XX 07-JUN-1995; 95US-00485243.

XX 16-JAN-1998; 98US-00007999.

XX 16-JAN-1998; 98US-00008172.

XX 20-JAN-1998; 98US-00009620.

XX 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes F1 or F2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilises the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36

XX SQ Sequence 1430 AA;

Query Match 43.7%; Score 3163; DB 5; Length 1430;

Best Local Similarity 47.2%; Pred. No. 2.4e-179;

Matches 675; Conservative 214; Mismatches 437; Indels 104; Gaps 38;

QY 1 MEKNLRYKLHKVKQWVAIGVT--VTLSFLAGGVVAAADTNNDGTSV----QVKNKV 53

Db 1 METKRYKMHKVKKHVTVAVASGLITLTGTLGSSVSAETEQTSKDVTKQKSEDDKAA 60

QY 54 PSDPKFDA---QAQNGQLAQMFKAAANOADTATTSQVSPATDGRVDNVTPAANQPAAN 109

Db 61 SESSQTDAPKTKQAQTEQ-TQAQ-SQANVAD-TSTIS---ITKETPSQNIITQANSDDKT 113

QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQDPG-----H 149

Db 114 VTNTKSEEAQTSEERTKQSEEAQTTSQAALTKQAKLTQRTAAQENKNPVDLAAIPN 173

QY 150 LETVDGKTYVDANGORLNKYSMWIDGKTYYPDGTGGEAQTDLP----KTGQANQNVDP 205

Db 174 VKQIDGKYIYIGSDQPKPKNFALTVNNKVLVFDKNTG-ALTDTSQYQFKQLTKLNN--- 229

QY 206 SYQANNQAYSNBASSFETVDNYLTADSWYRPRKILKNGQSWQASSEDRLPILMTWPPDA 265

Db 230 DYTPHNQIVNFENTISLETIDNVTDADSWYRPRKILKNGKTTATSESDLRPLIMSWPDK 289

QY 266 ATKAAYANFWAKEGLISG-SYRQNSA--NLDATQNIQSAIEKKIASEGNTWLRDKMSQ 322

Db 290 QTOIAVLNTWQOGLGTGENYTDADSSQESLNLAAQTQVKIETKISQTOQTQTLRLDIINS 349
QY 323 FVKSQNWSIASNETVYFNQDHMOGALLFSNSKDTTEHANSDWRLNLRNPTFQTGKQKY 382
Db 350 FVKTPNWSQTESDTSAGEKHLQGALLYSNSDKTAYANSYRLNLRNPTFSQTGKPKY 409
QY 383 FTTNYA-GYELLANDVNSNPVQAEQLNHLHLNMGWDIVMGDKDANPDGVRVDAVDN 441
Db 410 FEDNSSGGYDFLLANDIDNSNPVQAEQLNHLNMGWDIVMGDKDANPDGVRVDAVDN 469
QY 442 VNADLLQIQORDYYKAKYGTONEKNAIDHLSILEAWSGNDNDVVKDQNNFSLIDNDQRS 501
Db 470 VDADLLQIASDYLLKAYGVDSKSEKNAIHLNLSILEAWSNDNDPQYNKDTKGAQLPIDNKLRL 529
QY 502 GMLKAF-----GYASAYRGNLSNLATAGLNKRS-A-NPDSDPVPNYPYFIRAHDEVOT 552
Db 530 SLLYALTRPLEKADSKNEIRSGLEPVIITNSLNRSAEGKSERMANYIFIRAHDEVQ 589
QY 553 RTAKIIREKLGKTNADGLTNLTLDDLNKAFDINQDMNATDKVYYPNNLPMAYMWLQNK 612
Db 590 VIAKIKAQI-NPKTDGLT-FTLDELKQAFKIYNEDMRQAKKQYQTSNIPATAVALMSNK 647
QY 613 DTVTRVYGYDMYTDNGQYMATKTPFYNAIETLLKRIKIVAGGQ--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYSDDDQYMATKSPYYDAIDTLKARIKAYAGGQDMKITYVEGDKSHMDW 707
QY 665 S-SGILTSVRYGKANSASDAGNTETRNSGMALLNNRPNFRAYRN--LTLNNGAAHKSQ 721
Db 708 DYTGLVTSVRYGTGANEAATDQSEATKTQGMVITSNFSKLQNDKQVIVNNGAAHKNQ 767
QY 722 AYRPLLSTKDGATATYLNDSVDSDQYKYTDSQGNLSFSASELSQSVANAQVSGMTQVWP 781
Db 768 EYRPLLSTKDGATATYLNDSVDSDQYKYTDSQGNLSFSASELSQSVANAQVSGMTQVWP 826
QY 782 VCAADNODVTSPTSQATKQGNHLYHQSADALDSQVIEGFSNFQAFQASPDQITNAVIQX 841.
Db 827 VGASNDQDVRVAASNAKANATQVYESSSALDSQLIEGFSNFQDFVTKDSYDTNKKIAQN 886
QY 842 GDLFKSWGITOPEMAPQVYSSSDGTFLDSVILNGVAFSDRYDLAKSNKNKYSKODLANA 901
Db 887 VOLFKSWGITSFEMAPQVYSSSDGTFLDSVILNGVAFSDRYDLAKSNKNKYSKODLANA 946
QY 902 IKGLQAGIKVLDSLVPNQLYNLPKGVVTVATRVNQQYQAKSGATINKTPYVANTRSYG- 960
Db 947 VKALHKSGLQIADWVPDQIYNLPKGVVTVATRVNQQYQAKSGATINKTPYVANTRSYG- 1006
QY 961 DYQEQYGGKFLDQLKLPRLSTQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKTV 1020
Db 1007 DYQAKYGGAFSLAALKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVY 1066
QY 1021 LSE--GNKYLNLADGKFLPLTVLNTYQPOVSANGFISKNGIHYLDKNGQEVKNRP-K 1077
Db 1067 LKDNASDKYFELKGNQTYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
QY 1078 EISGSWYTFDSDGKMATKTKIGNDTYLPMPNGKQLKEGW--YDGKXAYYDDNGRTWT 1135
Db 1122 DAKGNWYTFDNGHMYGLQQLNGEVQYFLSNGVQLRESFLENADGSK-NYEGHLGNYS 1180
QY 1136 KNGFVEFRVGDQKRYFNGDGTIAIGLVSLNRLTYFDAYGYQVKGQVTV-INGKSYTF 1194
Db 1181 N-GYYSF--DNDSKRYFDASGMAVGLKTINGNTQYFDQGYQVKGAWITGSDGKRYF 1237
QY 1195 DADQGLVQTDNANPAPQOAGKLLGDNQWY-RKDGQLLTGQSTIDGQKFFPDNGVQ 1253
Db 1238 DGSNGMAVNPANDK-----NGDMWYLSNDGIALGVQTINGKTYFYQDQKQ 1286
QY 1254 VKGQGTATDASGLVRFYRDQGHQVKGWYSTDNNWYVNESGQVLTGLQITIDGQTVYFD 1313
Db 1287 IKGKIITD-NGKLYFLANSGLARNIFATDSQNNWYFGSDVAVTGSQITAGKLIYFA 1345
QY 1314 DKGIQAKGKAVMDENGLRYFDADSGNMLRDRWK-NVDGNWYFNRNGLA 1362

Db 1346 SDGKQVKSFV-TYNGKVHYHADSGELQVNRFEADKQGNWYLDNSGEA 1394
RESULT 11
ADD93658
ID ADD93658 standard; protein; 1554 AA.
XX ADD93658;
XX AC
XX 29-JAN-2004 (first entry)
XX Streptococcus sobrinus glucosyltransferase-U.
XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX Streptococcus sobrinus.
XX WO2003075845-A2.
XX 18-SEP-2003.
XX 07-MAR-2003; 2003WO-US006962.
XX 07-MAR-2002; 2002US-0363209P.
XX 08-AUG-2002; 2002US-0402483P.
XX (FORS-) FORSYTH INST.
XX Smith DJ, Taubman MA;
XX WPI; 2003-845091/78.
XX Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
XX Claim 16; Page 15; 49pp; English.
XX The present sequence is the protein sequence of Streptococcus sobrinus glucosyltransferase-U (GTF-U). Peptide fragments of GTF-U, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.
XX Sequence 1554 AA;
Query Match 42.6%; Score 3080.5; DB 7; Length 1554;
Best Local Similarity 43.9%; Pred. No. 2.3e-174;
Matches 667; Conservative 213; Mismatches 446; Indels 193; Gaps 34;
QY 1 MEKNLYKLYKLVKQWVAIGVTVTTLSPFLAGQVAA-----DTNNDG 44
Db 1 MEKLLYKLYKLVKQWVAIGVTVTTLSPFLAGQVAA-----DTNNDG 60
QY 45 TSVQVKNKVPSPDKFDAQONGQLAQMFKAAQADQ---TATQVSPATGVRVNDQVT 100
Db 61 TTNDANTTTTDTTADQASANTNQ-DQAGSDQSNNDQAKQDTANTDRNADNSQTDN---- 116
QY 101 FAANQPAANVANQDVANPATDAGALNRQSAADTST----DGKAVPQTSDPG-----HLE 151
Db 117 -----NQATDQATS PATDGTSGVQRDAANVATAADQEGQTASEQKSAALSNDVK 168
QY 152 TVDGKTYVDANGQLKRYKSWVIDKTYFDQGTGEAQTDLKPTQANQDNVDPVSYQANN 211
Db 169 LIDGKYTYVQADGSKYKGNFAITVNGQMLYFSDTGCALSSSTYSFSQGTTLNLDVDFSSH 228

Db 1 MEKKLHYKHVKHVVITIAVASIGLVSLVAGAGTVAEDKAVANDTTAQTAVGVDTGQDQA 60
QY 45 TSVQVNMVSPDPKEDAQAGQAGLAQAMFAKANQADQ-----TATSOVSPATGRVDNQVT 100
Db 61 TTNDANTNTTDDTADQASANTNQ-DQAGSDQSNQDQAKQDPTANTDRNQADNSQTDN--- 116
QY 101 PAANQPAANVQDVANPATDAGALNRQSAADTST-----DGKAVPQTSQDPG-----HLE 151
Db 117 -----NOATDQATSPATDGTSTVQRDAANVATAADQEGTAPSEQKSAALSIDNVK 168
QY 152 TVDGTYYVDANGORLKNYSWIDGKTYFFDQGTGEAQTDLPKTQANODNVPDSQANN 211
Db 169 LIDGKYVYQADSGYKKNFAITVNGMLYFDSDTGALSSTYSQGTNTLVDFSSH 228
QY 212 QAYSNEASSFETVDNYLTADSYRPRKILKNQSQWASSEGDLRPLMTWPDAAKAA 271
Db 229 KAYDSTAKSFELVNGVLTANSWYRPAIILRNQGTWEASNENLRLPVLMSWPDKDTQVAY 288
QY 272 ANFW-+-AKGLISGYSRONSANLDAATQNIQIAIEKIASEGTNWLRDNKQSFVKS 326
Db 289 VNYMKYLSANETEVNTNETSQVDLKEA--QSIQTKIEQKITSNDSQTLRTAMEAFVAA 346
QY 327 QNOWSIASNETVYPNQDHMOGALLFNSKDEHANSQWRLNBNPTFQTKQKYFTT- 385
Db 347 QPKWNNSTEN---FNKGDHLOGGALLYTNSDLTPWANSDYRLNLRPTTQOQGTKKYFTG 403
QY 386 NYAGYELLANDVNSPNVQAEQLNLHLYLNNWGDIVMGDKDANFDGVRVDAVDNVD 445
Db 404 GEGGYEFLLSNDVNSPNVQAEQLNLHLYLNNWGDIVMGDKDANFDGVRVDAVDNVD 463
QY 446 LLOIQDYKAKYGTQDQEKNAIDHLSILEAWSNDNDYVQDQNNFSLSIDNQSGMLK 505
Db 464 LLOQVSYNFQKQYKVTDSSEANALAHISILEAWSLNDQNYEDTNGTALSIDNSSRLTSLA 523
QY 506 AFGYASARCNLSNLTAGLKNRSNPDG--DPVNYVPIRAHDSQVTRIARIKREKL 563
Db 524 VLTKPGQRIIDLNSLISESVKNERANDTAYGDTIPTYSFVRAHDSQVTRIARIKREKI- 582
QY 564 KTNADGLNLTLDLNLKAFDIYNQDMNATDKVYYPNNIPMAVAMLONKDTVTRYVYGD 623
Db 583 DTNSDGYT-FILDQLKDAFKIYNEDMAKVNKYTHYNIPAAVALLSNWESVPRVYGD 641
QY 624 YTDNGQYMAKTPFPYNAIETLKGRIKYVAGQAVS-YKQDMSGILTSVRYGKANSAS 682
Db 642 YTDGQYMAKSPYDAIATMLQGRIAVYSGQSEEVHKVNGNQLSSVRYGQDLSAD 701
QY 683 DAGNTE--TNSGSMALLINRNP-RAVRLNLTLMCAHKSQYRPLLSTKDGATYND 740
Db 702 DTQGTDLSTGLVTLVNDPNDLIDGGPSLTVMNGRAHQAYRPLILGTGQVQSLKD 761
QY 741 SDVDSRQYKTDSONLSPSASELOSVANAQVSGMIQVWPVGAADNODVRTSPSTQATK 800
Db 762 SDTHI--VKYTDANGNLTFTADDIKGYSTVDMSGYLAVVWPVGAQDQDVRVAADNOKA 819
QY 801 DGNVYHSDALDSQVYIEGFSNFOAQPSPDQYTNVIAKNGDLKFSWGITQFEMAPQY 860
Db 820 DGKSLKTSAAALDSQVYIEGFSNFOFANNDADYTNKKAENADPFKLGITSEFEMAPQY 879
QY 861 SSEDGTFDLSVLNGYAFSDRYDLAMSKNKGKQDLANAIGKLOSAGIKVLSLDVNO 920
Db 880 SATDGSFLDSTIQNGYAFSDRYDLAMSKNKGKQDLANAIGKLOSAGIKVLSLDVNO 939
QY 921 LYNLPGEVTVATRVNQYQAKSGATINKTPVANTRSYG-DYOSQYGGKFLDLOKLYP 979
Db 940 IYQLPGEVTVATRVNQYQAKSGATINKTPVANTRSYG-DYOSQYGGKFLDLOKLYP 999
QY 980 RLFTSKJSTGKPIDPSVKITNWSAKYFNGSNILGKARYLSEG--NKYLMNLDAGKFL 1037
Db 1000 DMFTVMNISTGKPIDPSVKIKQWEAKYFNGSNILGKARYLSEG--NKYLMNLDAGKFL 1059
QY 1038 PTVLNTYQVQSVANGFISKNGGHIHYLDKNGQEVKNRPFKEISGWSYFFSDGKATG-- 1095
Db 1060 PA-----SPTGQDNKATGYFDGTGMAYYSTSGNKAIVNSFIVEGGHYHYFFDKDGHMVTG 1115

QY 1096 KTKIGNDTYLFMPNGKQLKEGVYDQK-KAYYYDDNRTWTNKG-----FVSPRVQDQK 1149
Db 1116 KAEDGND-YYFLPNGIQMRDAIYQDAQGNYYGRTGILY--KGDWYFPVDPNNANKTV 1172
QY 1150 WRYFNGDGTIALGLVSLNRLTYLPAAYGVQVKGQVTTNGKSYTFDADQG----- 1199
Db 1173 FRYFANNVMAIGYRNMYGQTYFFDENGFQAKQLLTDKGTHTYFDEDNAMAKNKFVNV 1232
QY 1200 --DLVQTNANPAPOG----- 1214
Db 1233 GDDWTYMDONGNAVAGQYFPVNNQILYFNPETGVQVKGQFITDAQGRTSYDANSALKSS 1292
QY 1215 -----AGMKLLGDN--QWGY--RKDGQLLTGEGTIDGQKVFPQ-DN 1250
Db 1293 GFFTPNGSDWYVAENGYVYKGFQVAENQDQWYFDQTTGKQAKGAQVDRDLYFNPDS 1352
QY 1251 GVQYKGGTATDASGVLRYFDRDQGHQVKGWYSTSDNNWYVNESGVULTGLQTD----- 1306
Db 1353 GVQYKGGTATDASGVLRYFDRDQGHQVKGWYSTSDNNWYVNESGVULTGLQTD----- 1412
QY 1307 -----GQTVYFD--DKGLOAKGKAVMDENGRLRYFDADSGNMLRDRWKN 1348
Db 1413 HFDEVTGQQAAGAAUVNGQQLYFVDVSGLOVKGDFVTDGQGTSTYDVNSGD-----KK 1466
QY 1349 VDG-----NWYFNRNG 1360
Db 1467 VNGFPTTGDNAWYADGOG 1485

RESULT 13

ADD93657

ID ADD93657 standard; protein; 1590 AA.

AC ADD93657;

DT 29-JAN-2004 (first entry)

DE Streptococcus sobrinus glucosyltransferase-I.

DX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

OS Streptococcus sobrinus.

PN WO2003075845-A2.

PD 18-SEP-2003.

PF 07-MAR-2003; 2003WO-US006962.

PR 07-MAR-2002; 2002US-0363209P.

PX 08-AUG-2002; 2002US-0402483P.

PY (FORS-) FORSYTH INST.

PI Smith DJ, Taubman MA;

PS WPI; 2003-845091/78.

PT Composition useful as vaccines for dental caries comprises a fragment of

PT a glucan binding protein-B binding to a major histocompatibility complex

PT class II protein.

PS Claim 16; Page 14; 49pp; English.

CC The present sequence is the protein sequence of Streptococcus sobrinus

CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially

CC from the catalytic domain of the polypeptide, can be used in immunogenic

CC compositions and subunit vaccines for dental caries. These compositions

CC comprise a major histocompatibility complex (MHC) class II protein-

CC binding peptide from S. mutants glucan binding protein-B (GbpB)

CC covalently linked with a peptide fragment of a streptococcal

CC glucosyltransferase. The compositions are used in a claimed method of

CC eliciting production of an antibody in a mammal. Diepitopic or
CC multiepitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
SQ Sequence 1590 AA;

Query Match 41.9%; Score 3031; DB 7; Length 1590;
Best Local Similarity 41.6%; Pred. No. 2.1e-171;
Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;

QY 1 MEKNRYKLHKVKKQVAGVITVTN-LGFLAGQVVAADTNNDGTSVQVNVKVPDPE 59
DB 1 MEKNRYKHKVKKQVAGVITVTN-LGFLAGQVVAADTNNDGTSVQVNVKVPDPE 55

QY 60 DAQAQNGQLAQAFAKAAQADQATATQSPATDGRVDNQVTPAANQPAANVANQDV- 115
DB 56 --QTNNQATDQTSIAA-----TATSEQASDTAATDQ--ASAAEQTQGTASTDTAAQT 106

QY 116 ----ANPA-----TDAGALNRQSAADTSTGKAVPQTSQDPQGHLETVDGKTYIV 160
DB 107 TTNAEAKVWPTENENQGFTEMLAEAKNVATAESD--SIPDLAKMSNVKQVDGKYY 164

QY 161 DANGQRLKNYSVIDGKTYVFDGQ--EAQTDLPKTQGANQDNVDPDYOANNOAYSN 216
DB 165 DQGNVKNFVAVSGDKIYFDF-ETGAYKDTSDVADKSSASVQNA-TIFAANNRAYST 222

QY 217 EASSPETYDNYLTADSWRPRKILKNGOSQWASSEGDLRPLMTWPDAAAKAAYANFWA 276
DB 223 SAKNEAVDNYLTADSWRPRKILKNGOSQWASSEGDLRPLMTWPDAAAKAAYANFWA 282

QY 277 KEGLSGSY--RONSANLDAATONTQSAIEKKIASEGNTNMLRDKOSQFVKSONQWSTAS 334
DB 283 KVGIDKTYTAETSQADLTAAAEVQARIEQKITSENNTKMLREAISAFVKTQPOWNGES 342

QY 335 ENETVYNQDHWOGALLFNSKD--TEHANSWRLLNBNPTQCK-QKYFTTN---YA 388
DB 343 EK----PYDDHLQNGALLFDNQTDTPTQSNRYLLNRTPTNQTGSLDSRFTYNDPLG 398

QY 389 GYELLANDVNSNPVQAEQLNHLHYLMNMWDGIVMGDKDANFDGVRVDADVNNADLLQ 448
DB 399 GYDFLLANDVNSNPVQAEQLNHLHYLMNMWDGIVMGDKDANFDGVRVDADVNNADLLQ 458

QY 449 IQDYVYKAKYGTQNEKNAIDHLSLEAWSGNDNDYVKDQNNFSLISINDQSRGMLKAPG 508
DB 459 ISSDYLKAAAGIDKKNKANNHVSIVEASDNDTPTLHDDGDNLMNDNKFRLSLWSLA 518

QY 509 YASAVRGNSLNATAGLNRSANP--DSDPVNPNVYFIRAHDSSEVOTRIAKITREKLGKNA 567
DB 519 KPLDKRSGNLNLSLVDREVDREVETVPSYSPARAHDSSEVDIIRDIHKAET-NPNS 577

QY 568 DGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAYAMWMLQNKDITVRVYGDYMTDN 627
DB 578 FGYSFTQEEIQAFKIYNEDLKKTDKXTHYNVPLSYLLTNKSGISPRVYVYGMFTDD 636

QY 628 GQYMATKTFYNAIETLLKRIKYVAGQVSKYQDWSGILTSVRYKSGANSASDAGNT 687
DB 637 GQYMAKNTVYDAIESLLKARMKYVSGGQAMQNYQIGNGEILTSVRYKSGALKQSDKGA 696

QY 688 ETRNSGMAILLNNRNFRA--YRNLTLNMGAAHKSQAYRPLLLSTKDGITATYLNDSVD-S 745
DB 697 TTRTSGVGVVMGQNFSLDQGVKVALNMGAAHANQYRALMVSTKDGIVATYLDADASKA 756

QY 746 ROYKYTDSQGNLSFSAEISQVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKDNIT 805
DB 757 GLVKRTDENGILYFLNDDLLKGVANQVSGFLQVWVPVGAADNQDIRVAASDTASTDGSL 816

QY 806 HQSDALDSQVIYEGSFNFOAFQSPDQYTNNAIKNKGLIFKSWGITQFEMAPQYVSSBDG 865
DB 817 HQDAAMDNRVMEFGSFNFOAFQSPDQYTNNAIKNKGLIFKSWGITQFEMAPQYVSSBDG 876

QY 866 TFLDSVILNGYAFSDRYDLAMSKNNKYSKQDLANAIKGLQSAKIKVLSLDLPNOLYNLP 925
DB 866 TFLDSVILNGYAFSDRYDLAMSKNNKYSKQDLANAIKGLQSAKIKVLSLDLPNOLYNLP 925

DB 877 QFLDSVIONGYAFTDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADVDPQMYTFP 936
QY 926 GKVVVATRVNQYQAKSGATINTKPYVANTRSYG-DYQEOYGGKFLDDLOKLYPRLPST 984
DB 937 KQEVVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDLDELKKEYPLELTK 996
QY 985 KOISTGKPIDPSVKITNWSAKYFNGSNLGRGAKYVLSE--GNKYNLNLADGKLPLPTVLN 1042
DB 997 KOISTGQAIQPSVKIKOMSAKYFNGSNLGRGADYVLSDQVSNKYFNVASDTLFLPSSL 1056
QY 1043 NTYGQPOVSANGFISKNGGIIHYLDKNGOEVRKRRFKEISGSWYFSDGKMATGKTIGND 1102
DB 1057 GKVESGIRYDQ---KGYIYNSSATGDQVAKASFITEAGNLNLYFGKDCGYMTGATINGA 1112
QY 1103 TYLFWPENGKQLKEGQVWVDGK-KAYYDDNGRTWTKNKFVPRVDGQDKRWYFNGGDTIAI 1161
DB 1113 NYFFLENGTALRNTIYTDQGNSHYANDGRYEN-GYQGF---GND-WRYFK-DGNMAV 1166
QY 1162 GLVSLD----- 1167
DB 1167 GLTTVDGNGVYFDKDGVOAKDKIIIVTRDGKVRVFDQHNNGNAATNTFIADTGHWYVLGD 1226
QY 1168 -----NRTLYFDAYGYQVKGQTVTIN-GKSYTFDADQGD----- 1201
DB 1227 GVAVTGAQTVGKQKLYFEANGQVKGDFVTSDEGKLYFYDVSQDMWTDTFIEDKAGNWF 1286
QY 1202 ----- 1201
DB 1287 YLKGDKAAVTGAQTIRGQKLYFKANGQVKGDIIVKGTGKIRYYDAKSQGEQVFNKTVKAA 1346
QY 1202 -----VOTDNANPAP----- 1231
DB 1347 DKTYVINGDGVADVDPSPVVKQGTQKADSGALRFLYNLKGQVLTGSGWYETAHNDWYIOSG 1406
QY 1232 QLLTGEQITDQKQVFFQDNGVQVKGATDASGVLRFYDRQGHQ----- 1276
DB 1407 KALIGEQTINGQHLVFKEDGHQVKGQVLTGTDGKVRYYDANSQDAFNSKSVTVNGKTYFF 1466
QY 1277 -----VKGWYSTSDDDNWVYVNSGQVLTGLQ 1303
DB 1467 GNDGTAQTAGNPKGTQKDGSDIRFYSMEGQVLTGSGWYENAGQWLYV-KNGKVLTLGLQ 1525
QY 1304 TIDGQTVYFDDKQLOAKGKAVDENGRLRYFDADSGNMLRDRKQNVKNDGNYWYFNENG 1360
DB 1526 TVGSRVYFDENGIQAKGKAVRTSDGKIRYFDENSGSMITNQWKFVYGOYYFFGNDG 1582

RESULT 14
ADX37280
ID ADX37280 standard; protein; 1590 AA.
XX
AC ADX37280;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus sobrinus glucan binding protein B.
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX Streptococcus sobrinus.
XX US2005031633-Al.
XX 10-FEB-2005.
XX 09-MAR-2004; 2004US-00797821.
XX 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-0029004P.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.


```

PR 07-MAR-2003; 2b03US-00383930.
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
DR
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 7; SEQ ID NO 37; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
XX binding protein-B (GbpB) and a biocompatible microparticle, where the
XX fragment binds to a major histocompatibility complex (MHC) class II
XX protein. The composition is useful for producing an antibody for
XX immunizing mammals against dental caries. This sequence corresponds to a
XX Streptococcus sobrinus GbpB protein of the invention.
XX
XX Sequence 1590 AA;
XX
XX Query Match 41.9%; Score 3031; DB 9; Length 1590;
XX Best Local Similarity 41.6%; Pred. No. 2.1e-171;
XX Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;
XX
QY 1 MEKURYLKHKVKKQWVAIGVTVT-LFSLAGQVVAADTNNDGTSVQVNRKQVSPDKPF 59
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MEKURYFKMHKVKRVTLSVASATMLASALGASVASADTDTASDDSNQA--VVTGD--- 55
QY 60 DAQAOQGLAQAFKAAQADQATATSOVSPATDGRVDNQVTPAANQPAANQVADV--- 115
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 --QTNQATQDTSIAA-----TATSEQASASTDAATDQ--ASAEQTCGTASTTAAQT 106
QY 116 ---ANPA-----TDAGALNRQSAADTSTDGKAVPQTSQPGHLETVDCGKTYV 160
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 TTNAEAKVPTENENQGFDEMLAEAKNVATAESD--SIPSDLAKMNVKQVQDKYV 164
QY 161 DANGORLKNYSVIDGKYVDFGQTG----EAQTLPTKGQANQNDVPSVQANQAYSN 216
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 DQDGNVKNFVAVSGDKIYFED-ETGAYKDTSKVDADKSSSAVSQNA-TIFAAANNRAYST 222
QY 217 BASSPETVDNYLTADSWYRPRKILKNGQSQWASSEGDLRPILMTWMPDAAKAAANFWA 276
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 SAKNFEADVNYLTADSWYRPRKILKNGQSQWASSEGDLRPILMTWMPDAAKAAANFWA 282
QY 277 KEGLTSGSY--RQNSANLDAATQNTQSAIEKKIASGNTWMLRDKMSQFVKSONQWSIAS 334
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 KVGIDKTYTAETSDADLTAAELVQARIEQKITSENTWKULREAISAFVKTPQWNGES 342
QY 335 ENETYPNQDHQGGALLFSNSKD--TEHANSDWRLNLRNPTFQTKG-QKYFTTN----YA 388
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 EK---PYDDLQNGALLFDNQTLTPTDQSNYRLNLRNPTNQTSGLDSRFTYNDPLG 398
QY 389 GYELLANDVNSNPVQAEQLNHLHLYLWNGDIIWNGDKDANFDGVRVADVNDVNDLQ 448
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
399 GYDFLLANDVNSNPVQAEQLNHLHLYLWNGDIIWNGDKDANFDGVRVADVNDVNDLQ 458
QY 449 IORDYKAKYGTQDQEKNAIDHLSLTLEAWSGNDNDYVQDQNNFSLSDNDORSGLKAFG 508
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
459 ISSDYLKAAYGIDKKNKANNHVSIVEAWSNDTPTLHDDGDNLMNDKFKLSLWLSIA 518
QY 509 YASAYRGNLSNLATAGLKNRANP--DSDPVNPNVYFIRAHDSVQTRIAKIREKLKGTNA 567
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
519 KPLDKRSGLNPLIHNLSLVDREVDDREVETVPSYSFARAHDSVQDIIRDIIKAEI-NFNS 577
QY 568 DGLTNLTLDLNLKAFDPIINQDNKATDKTYVNNLPMAYAMLQNKDVTTRVITYGDMYTDN 627
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
578 FGY-SFTQEEIEQAFKIYNEDLKKDKKYTHYVPLSYTLTLTNKSGIPRVYVYGMDFDD 636

```

RESULT 15

AAR91047

ID AAR91047 standard; protein; 1577 AA.

XX AAR91047;

XX 16-OCT-2003 (revised)

DT 22-MAY-1996 (first entry)

```

QY 628 GQYMATKTPFYNAIETLLKGRIKYVAGGQAVSYKQDWSSGILTSVRYGKAGNASDAGNT 687
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
637 GQYMAKNTVNYDAIESLLKARMKYVSGGQAMQNYQIGNGEILTSVRYGKALKQSDKQGA 696
QY 688 ETRNSGMALLNNRNPFRA-YRNLTLNMGAAHQAAYRPLLLSTKQGINIYTLNDSVD-S 745
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
697 TTRTSGVGVVMGNQNFSLDGKVALNMGAAHQAAYRPLLLSTKQGINIYTLNDSVD-S 756
QY 746 RQYKYTDSQGNLSFSAEQLSVANAQVSGMIQVVPVGAADNQDVRTSPTQATKQDNLY 805
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
757 GLVKKTDENGLYFLNDDLKGVANQVSGFLQVWPVGAADQDQIRVAASDQSTASTDGKSL 816
QY 806 HQSDALDSQVIYEGFSNFQAFQSPDQYTNVIAKNGDLFKSWGITQFEMAFQYVSSSDG 865
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
817 HQDAAMDSCRVMPEGSFNFQSPATKEEYTNVVIANNVDKFSVSGITDFEMAQYVSSD 876
QY 866 TPLDSVILINGAFSRYDLAMSKNNKYSKQDLANAIGLQAGIKVLSDLYPNQLYNP 925
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
877 QPLDSVIONQYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFP 936
QY 936 GKEVVTATRVNQYQAKSGATINKTPYVANTRSYG-DYQEOYGGKFLDLDLQKLYPRLFST 984
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
937 KQEVVTATRVNQYQAKSGATINKTPYVANTRSYG-DYQEOYGGKFLDLDLQKLYPRLFST 996
QY 985 KQISTGKPIDPSVKITNMSAKYFNGSNILGRGAKTVLSE--GNKYLNLDLGLFPTVLN 1042
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
997 KQISTGKPIDPSVKITNMSAKYFNGSNILGRGAKTVLSE--GNKYLNLDLGLFPTVLN 1056
QY 1043 NYQGOQVQVANGPTISKNGGIHYLDKNGQEVKVRPKBISGSWYVPSDQMGKATGKTKIGND 1102
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1057 GKVESGIRYDG---KGYIYNSSATGQVQKASFTIAGNLVYFQKQVWVYGAQTINGA 1112
QY 1103 TYLEMPNCKQLKEGVVYDQK-KAYYDDNGRTWTKGFEVERVDCQDKWRYENGDTIAI 1161
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1113 NYFLENGTALRNTIYTDQGNSHYANDGKRYEN-GYQGF---GND-WRYPK-DGNMAV 1166
QY 1162 GLVSLD----- 1167
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1167 GLTVDGNVYFQKQVQKDKIIVTRDGKVRYPQHNNGNAATNTFIADKTHWYVYLGKD 1226
QY 1168 -----NRTLYFDAYGYQVKGQVTTIN-GKSYTFDADQGL----- 1201
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1227 GVAVTGAQTVGKOKLYFEANGQVQKGFVTSDEGLKLYFYDVGSDGMDMTDTFIEDKAGNWF 1286
QY 1202 ----- 1201
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1287 YLKGKGAAVTGAQTIROKLYFKANGQVQKGDIVKGTGKIRYDQKSGEQVFNKTVKAA 1346
QY 1202 -----VQDNDANPAP-----QGG-----AGWKLKGNQWGYRKDG 1231
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1347 DGKTYVIGNDGVAVDPSVVKQGTQKASGALRFFYNLKGQLVTGSGWYETAHHDWYVYISG 1406
QY 1232 QLLTGEQITDQKVFQDQNGVQVKGATATDASGVLRFYDRDQGHQ----- 1276
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1407 KALATGEQITINGHLYFKEDGHQVKGQLVTGTDGKVRYYDANSQDQAFNKSVTNGKTYVF 1466
QY 1277 -----VKGHWYSTDDNNVYVNSGQVLTGLQ 1303
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1467 GNDGTAQTAGNPKQGTQKGSDDIRFYSMEGQLVTGSGWYENAAQGGMLYV-KNGKVLTLGLQ 1525
QY 1304 TIDGQTVFDDKGIQAKGKAVDENGRLRYFDADSGNMLRDRKWKVDGNYWYFNRRNG 1360
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1526 TVGSRVYFDENGIQAKGKAVRTSDGKIRYFDENSGSMITNQWKFVYQYVYFGNDG 1582

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 19:39:30 ; Search time 29.8857 Seconds
(without alignments)
3776.130 Million cell updates/sec
Title: US-10-797-821-39
Perfect score: 7230
Sequence: 1 MEKNLRYKLHKYKQWVAIG.....MKNVNDNYYFNRNGLATRW 1365
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 572060 seqs, 82675679 residues
Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Issued Patents_AA*
1: /cgn2_6/ptodata/1/1aa/5 COMB.pdp.*
2: /cgn2_6/ptodata/1/1aa/6 COMB.pdp.*
3: /cgn2_6/ptodata/1/1aa/H COMB.pdp.*
4: /cgn2_6/ptodata/1/1aa/pCTUS COMB.pdp.*
5: /cgn2_6/ptodata/1/1aa/RE COMB.pdp.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pdp.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.
SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3174	43.9	1430	2	US-09-008-172-2 Sequence 2, Appli
2	3174	43.9	1430	2	US-09-210-361-6 Sequence 6, Appli
3	3174	43.9	1430	2	US-09-740-274-6 Sequence 6, Appli
4	3026.5	41.9	1577	1	US-08-793-824-2 Sequence 2, Appli
5	2945.5	40.7	1475	2	US-09-007-999-2 Sequence 2, Appli
6	2945.5	40.7	1475	2	US-09-210-361-2 Sequence 2, Appli
7	2945.5	40.7	1475	2	US-09-740-274-2 Sequence 2, Appli
8	2893.5	40.0	1375	2	US-09-210-361-4 Sequence 4, Appli
9	2893.5	40.0	1375	2	US-09-740-274-4 Sequence 4, Appli
10	2227	30.8	2057	2	US-09-499-203-2 Sequence 2, Appli
11	2191.5	30.3	1278	2	US-09-604-957-3 Sequence 3, Appli
12	2191.5	30.3	1781	2	US-09-995-749A-2 Sequence 2, Appli
13	1635.5	22.3	545	2	US-09-604-957-4 Sequence 4, Appli
14	1597.5	22.1	545	2	US-09-995-749A-10 Sequence 10, Appli
15	1362.5	18.8	523	2	US-09-604-957-5 Sequence 5, Appli
16	1361	18.8	522	2	US-09-995-749A-11 Sequence 11, Appli
17	1203.5	16.6	535	2	US-09-604-957-7 Sequence 7, Appli
18	1203.5	16.6	535	2	US-09-995-749A-13 Sequence 13, Appli
19	1173	16.2	584	2	US-09-604-957-6 Sequence 6, Appli
20	1171	16.2	584	2	US-09-995-749A-12 Sequence 12, Appli
21	450	6.2	349	2	US-09-009-620-2 Sequence 2, Appli
22	365.5	5.1	2710	1	US-08-480-604A-6 Sequence 6, Appli
23	365.5	5.1	2710	1	US-08-405-496A-6 Sequence 6, Appli
24	365.5	5.1	2710	2	US-08-915-136-6 Sequence 6, Appli
25	365.5	5.1	2710	2	US-08-957-310-6 Sequence 6, Appli
26	365.5	5.1	2710	2	US-10-011-366-6 Sequence 6, Appli
27	365.5	5.1	2710	2	US-09-084-517-6 Sequence 6, Appli

28	330.5	4.6	1231	2	US-08-714-741-41 Sequence 41, Appli
29	327.5	4.5	2366	1	US-08-480-604A-10 Sequence 10, Appli
30	327.5	4.5	2366	1	US-08-405-496A-10 Sequence 10, Appli
31	327.5	4.5	2366	2	US-08-915-136-10 Sequence 10, Appli
32	327.5	4.5	2366	2	US-08-957-310-10 Sequence 10, Appli
33	327.5	4.5	2366	2	US-10-011-366-10 Sequence 10, Appli
34	327.5	4.5	2366	2	US-09-084-517-10 Sequence 2, Appli
35	323	4.5	866	2	US-09-545-773-2 Sequence 2, Appli
36	323	4.5	866	2	US-10-222-038-2 Sequence 2, Appli
37	318.5	4.4	811	1	US-08-480-604A-7 Sequence 7, Appli
38	318.5	4.4	811	1	US-08-405-496A-7 Sequence 7, Appli
39	318.5	4.4	811	2	US-08-915-136-7 Sequence 7, Appli
40	318.5	4.4	811	2	US-08-957-310-7 Sequence 7, Appli
41	318.5	4.4	811	2	US-10-011-366-7 Sequence 7, Appli
42	318.5	4.4	811	2	US-09-084-517-7 Sequence 29, Appli
43	318.5	4.4	812	1	US-08-480-604A-29 Sequence 29, Appli
44	318.5	4.4	812	2	US-08-915-136-29 Sequence 29, Appli
45	318.5	4.4	812	2	US-09-084-517-29 Sequence 29, Appli

ALIGNMENTS

RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match	43.9%	Score 3174;	DB 2;	Length 1430;
Best Local Similarity	47.3%	Pred. No. 5.2e-217;		
Matches	677;	Conservative 213;	Mismatches 436;	Indels 104; Gaps 38;
QY	1	MEKNLRYKLHKYKQWVAIGVTT--VTLSFLAGGQVVAADTNNDGTSV-----QVNMKV	53	
Db	1	METKERYKMHKVKHVVTVAVASGLITLGTTLGSSVSAETEQOTSQSDKVVTKSEDDKAA	60	
QY	54	PSDPKFDA-----QAQNGQLAAMFAKNAQDQTATQSVSPATDGRVDNQVTPAANQPAAN	109	
Db	61	SESSQTDAPTKQAQTEQ-TQAQ-SQANVAD-TSTS-----ITKETPSQNIITTAQNSDDKT	113	
QY	110	VANDVANPATDAGALNFQSAADTSDCKAVP-----QTSDPG-----H	149	
Db	114	VTNTKSEEAQTESEERTKQSEEAQTASSQALTKQAEUTKQRTAAQENKPNVDAALIPN	173	
QY	150	LETVDGKTYTYDANGQRLKNTSMVIDGTYTFDGTGEAQTDLP-----KTQANDQNDVPD	205	
Db	174	VKQIDGKYYIIGSDGQPKKFNALTNNKVLFDKNTG-ALTDTSQYQFQGLTKLNN----	229	
QY	206	SYQANNQAYSNEASFETVDNYLTADSWYRPRKILKNGSQWASSEGDLRPLMTWPDPA	265	
Db	230	DYTPHNQIVNFENTSLKETIDNYVTADSWYRPRKILKNGKTWTASSESDLRPLLMWSWPK	289	
QY	266	ATKAYANFWAKEGLISG-SYRQNSA--NLDAANTQNIQSAIEKKIASSEGTNTWLRDKMSQ	322	
Db	290	QTQIAYLNNMQQGLGTGENTADSSQSLNLAQTVQVKIETKISQTOQTQWLRLDIINS	349	
QY	323	FVKSQNWSIASENETVYPNQDHMQGALLFSNSKDTTEHANSWELLNLRNPTFTQTKQKY	382	

Db	350	FVKTQPNWNSQTESDTSAGSKOHLQGGALLYNSDKTAYANSYRLLNRTPTTSQTKPKY	409
Qy	383	FTTNYA--GYELLALLDVNSPPVQABQLNHLHYLMNWGDI VMGDXDANFDGVRVDAVDN	441
Db	410	FEDNSGGYDFLLANDIDNSPPVQABQLNHLHYLMNYSIVANDPEANFDGVRVDAVDN	469
Qy	442	VNADLLOIQIRDYYKAKYGTDOEKNAIDHLISLEAMSGNDNDYVKQNNFSLIDNDQRS	501
Db	470	VNADLLQIASDYLKAHYGVDPSEKNAIHLISLEAMSGNDNDPQYNKDTKGAQLPIDNKLRL	529
Qy	502	GMLKAF-----GYASAVRGNLSNLTAGLAKNSA--NPDSPDPVNPYPIRAHDSVQT	552
Db	530	SLLYALTRPLEKQASNKQEIIRSGLEPVITNSLNRSAEKGKNSRMANYPPIRAHDSVQT	589
Qy	553	RIAKIIRKLGKTNWADGLTNLLTLDLNLKAFDIYNQDMNATDKVYPYNNLPMAYMWLQNK	612
Db	590	VIAKIIKAQI--NPKTDGLT--FTLDELKQAFKIYNEWMRQAKKXYTOSNIPTATALMLSNK	647
Qy	613	DTVTRVYVGYMDYNDGOYMAKTPFPYNAIFTLKGRIKYVAGQ--AVSYKQ-----DW	664
Db	648	DSITRLYVGYMDYSDGGYMAKTPPYDAIDTLLKARIKYAAGGQDMKITYVEGDKSHMDW	707
Qy	665	S--SGILTSVBYGKGANSASDAGNTETETNSGMAILNRPFRAYRN--LTLANGAAHKSQ	721
Db	708	DYTGVLTSVRYGTGANEATDQGEATKTQGMVITSNPSLKLQNQNDKVIYNNGAAHKNQ	767
Qy	722	AYRPLLSTKDGITATYLDSDVDSDROYKYTDSQGNLSFSASLSQSVANAAQVSGMIQVWVP	781
Db	768	EYRPLLTTKDGLTYSYSDAAKS--LYRKTNDKGELVFDASDIQGYLNPQVSGYLAVWVP	826
Qy	782	VGAADNQDVRTSPSTQATKQGNITYHOSDALDSQVIVYEGFSNFQAFQSPQDYTNVIAKN	841
Db	827	VGASDNQDVRAASNAKANATQVYESSALDSQLIYEGFSNFQDFTVKDSDYTNKXIAQN	886
Qy	842	GDLFKSGIITQFEMAPQVYSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYGSKQDLANA	901
Db	887	VQLFKSGVGTFSFEMAPQVYSSDGSFLDSITQNGYAFEDRYDLAMSKNNKYGSGQDMINA	946
Qy	902	IKGLQSAGIKVLSDLVPLNQLYNI PGKEVVTATRVNOYGOAKSGATTNKTPYVANTRSYG-	960
Db	947	VKALHKSGIQVIADWVPDQIYNIPGKEVVTATRVNDYGEYRKDSEIKWTLIYAANTKNGK	1006
Qy	961	DYQEQYGGKFLDLOKLYPLRFLSTQIISTOKPIDPSVKIITWSAKYPNGSNILGRGAKYV	1020
Db	1007	DYQAKYGGAFSLSELAAYPSIFNRTQISNGKKIDPSEKITA WKAKYFNGTNILGRGVYV	1066
Qy	1021	LSB--GNKYLNLADGKLFLEPTVLNNTYGPQVSVANGFISKNGGIHYLDKNGQEVKNRF-K	1077
Db	1067	LKONASDKYFELKGNQTYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ	1121
Qy	1078	EISGSWYTFPDSDGKMATGKTIGINDTYLFPNGPKLKEGVW--YDGGKAYYYDDNGRWT	1135
Db	1122	DAKGNWYTFDNGHWYVGLQQLNGEVQYFSLNGVQLRESFLENADGSK-NYFGLHGNRYS	1180
Qy	1136	NKGFPVFRVDGQKWRVYFNGDGTIAIGLSLDNRTLFDYAYGYQVKGQTVT--INGKSYTF	1194
Db	1181	N-GYYSF--DNDSKWRVFDASGVMAVGLKTINGNTQYFDQDQGYQVKGAWITGSDGKKRYF	1237
Qy	1195	DAQGDVLVQTDNANPAPQOGAKWKLGDNONGY--RKDGQLITGEQITDQKVFQDNGVQ	1253
Db	1238	DCSGNNVANRFPANDK-----NGDWYLYNSDGIALVGQTINGKTYTFYFGDGKQ	1286
Qy	1254	VKGGTATDASGVLRFFYDRDQCHGVKGWYSTSDDNWYVYVNESGOVLTLGLQTDGQTVYFD	1313
Db	1287	IKGKIITD--NGKUKYFLANSGBELARNIFATDSQNNWYFQSDGVAVTGSQTIAGKKLYFA	1345
Qy	1314	DKGIOAKGKAVBDENGNLRYFDADSGNMLRDRWK-NVDGNWYVFNRRGLA	1362
Db	1346	SDGKQVKGVSFV--TYNGKVHYHADSGBSLOVNRFEADKGNWYLYDSNGEA	1394

```

US-09-210-361--6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361--6

Query Match          43.9%; Score 3174; DB 2; Length 1430;
Best Local Similarity 47.3%; Pred. No. 5.2e-217;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

Qy      1  MEKRLRYKLHKVKKQVAGVTT--VTLSFLAGGVAAADTNNDGTSV-----QYNKMV 53
Db      1  METKRRYKHKHKVKKHWTVAVASGLITLGTTLTGSSVSAETEQOTSQDKVVTQKSEDDKAA 60

Qy      54  PSDPKFDA---QAQNGOLAQMFKAAQADOTATSQVSPATDGRVDNVOVTPAANQPAAN 109
Db      61  SSSQTDAPKTKQAQTEQ-TQAQ-SQANVAD-TSTS-----ITKETPSQNTTQANSDDKT 113

Qy      110  VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDPQG-----H 149
Db      114  VTNKSEEAQTSEERTKQSEEAQTTASSQAALTQAKAELTKQRQTAQENKNPVDLAAIPN 173

Qy      150  LETVDGKTYIVDANGRLKKNYSWIDGKTYYPFGQTGEAQTDLP-----KTGQANQDNVPD 205
Db      174  VKQIDGKYYIGSDGPGKKNFALTVNKVLDFDKNTG-ALTDTSTSYQYQFKQGLTKLAN--- 229

Qy      206  SYQANQVAYNEASSFETVDNYLTADSWRPRKILKNGOSQWASSEGDLPILMTWMPDA 265
Db      230  DYTPTHQIVNFENSTLETIIDNYTADSWRPRKOLLKNGKWTWTASSSDRLPLMSWWPDK 289

Qy      266  ATKAAYANFWAKEGLISG-SYRONSA--NLDAATQNIQSAIEKKIASSEGNTNWLRDKMSQ 322
Db      290  QTQIAYILNYNQOGLGTGENYTADSSQESINLAAQTQVVKIETKISQTQQTQMLRDIINS 349

Qy      323  FVKSQNWQSTASNETVYPNQDMQGALLFNSNKOTEHANSQWRLNLRNPTTQTGKQKY 382
Db      350  FVKTQPNWNSQTSDESISAGEKDHLOGGALLYSNSDKTAYANSIDSYRLNLRNPTTQTGPKPY 409

Qy      383  FTTNYA-GYELLANDVDNSNPVQAEQLNHLHVMNWDGIVMGDKDANFDGVRVDAVDN 441
Db      410  FEDNSGGYDFLLANDIDNSNPVQAEQLNHLHVMNYSIVANDPEANFDGVRVDAVDN 469

Qy      442  VNADLLQIQRDYKAKYGTQDNEKNAIDHLSILLEAWSGNDNDYVKQDNFSLSDINDQRS 501
Db      470  VNADLLQIASDYLKAHYGVQDKSEKNAIHLHSILLEAWSNDNDPOYNKDTKGQALPDKRLRL 529

Qy      502  GMLKAF-----GYASAVRGNLSNLATAGLKNRGA-NPDSQDPVNPVYFIRAHDSVQVT 552
Db      530  SLLYALTRPEKDAKSNKEIRSGLEPVTINSLNRSAGKNSRMANIYIFRAHDSVQVT 589

```



```
QY 553 RIAKIIRKLGKTNAGLNLTLDDLNKAFDIYNQDMNATDKVYPPNNLPMAYAWMLQNK 612
Db 590 VIAKIIKAQI-NPKTDGLT-FTLDELKQAFKIYNEQDMRQAKKKYQTSNIPTAYALMSNK 647
QY 613 DTVTRVYVYGDMDYDNGQYMATKTPFYNAETLLKGRIKYVAGGO--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYSDGQYMATKSPYDAIDTLKARIKAYAGGQDMKITYVGGDKSHMDW 707
QY 665 S-SGILTSVRYGKANSASDAGNTETNSGMALLNNRPNFRAYRN--LTLNNGAAHKSQ 721
Db 708 DYTGLVTSVRYGTGANEATDQSEATKTQGMVAVITSNPDLKQNDKVIYNMGAHKNQ 767
QY 722 AYRPLLSTKOGIATYLNDSVDVSQYKYVTDSCGNLSFASBLQSVANAQVSGMIQVWYP 781
Db 768 EYRPLLSTKOGIATYLNDSVDVSQYKYVTDSCGNLSFASBLQSVANAQVSGMIQVWYP 826
QY 782 VGADNDQVTRTSPQATKQGNHYHQSDALDSQVLYEGFSFOAQAQSPDQYTNVIAKQ 841
Db 827 VGASNDQVTRVVAASNAKATQVYESSALDSQVLYEGFSFOAQAQSPDQYTNVIAKQ 886
QY 842 GDLFKSGWITQEPMAPOVYSSBDGTFLOSVILNGVAFSDRYDLAMSNNKYGSKODLANA 901
Db 887 VQLFKSGWITQEPMAPOVYSSBDGTFLOSVILNGVAFSDRYDLAMSNNKYGSKODLANA 946
QY 902 IKGLQSAKIVLSDLPVNLQYVATRVNQQYQAKSGATINKTPYVANTRSYG- 960
Db 947 VKALHSGIQTIVADWPQIYNLPKGVVATRVNDYGEYRDSBKNTLYAANTKSNKG 1006
QY 961 DYQEOYGGKFLDLOKLYPRLPSTKQISTGKIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 1007 DYQAKYGGAFLELAAPYPSIFNRTQISNGKKIDPSKITAUKAKYFNGSNILGRGAKYV 1066
QY 1021 LSE--GNKYLNLADGKLFPTVLNNYQPOVQVANGFISKNGIHYLDKNGQEVKNRF-K 1077
Db 1067 LKDNASDKYFELKGNQVILPKQNTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
QY 1078 EISGSWYFYDSDGKMATKTKIGNDTYLFMPNGKQKKEGVW--YDGKAYYYDDNGRTWT 1135
Db 1122 DAKGNWYFYDNGHMYGLQQLNGEVQYFVLSNGVQLRESFLENADGSK-NYFGLHNGYS 1180
QY 1136 NKGFEVFRVGDQKRYFNGDGTIAIGLVSLDNRTLYFDAYQYKVGQTVT--INGKSYTF 1194
Db 1181 N-GYTSF--DNDSKRYFVDSAGVMAVGLKTINGNTQYFDQDQYQVKGAWITGSDGKRYF 1237
QY 1195 DADQGLVQTDNANPAPQOAGKLLGNQWGY-RKDGQLTGEOTIDGQKVFQDNGVQ 1253
Db 1238 DQSGMVAWNPANDK-----NGDWTYLSNGDIALVGVQTINGKTYFYQDQKQ 1286
QY 1254 VKGFTATDASGLYFRDQGHQVKGWYSTSDNNWYVNESGQVLTGLQITIDGQTVYFD 1313
Db 1287 IKGKIITD-NGKLYFLANSBELARNIFATDSQNNWYFGSDGVAVTGSQTIAGKLIYA 1345
QY 1314 DKGOAKKAVWDENGNRLRYFDADSGNMLRDRWK-NVDGNWYFNRNGLA 1362
Db 1346 SDGQVKGFSV--TYNGKVHYHADSGELQVNRFEADKQGNWYLYDSNGEA 1394
```

RESULT 3

```
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-Containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
```

```
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6
```

```
Query Match 43.9%; Score 3174; DB 2; Length 1430;
Best Local Similarity 47.3%; Pred. No. 5.2e-217;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;
```

```
QY 1 MEKNRYKHLKVKQWVAIGVTT--VTLSFLAGGVVAADTNNNDGTSV-----QVKNV 53
Db 1 METKRYKHKVKCHWTVAVASGLITLCTTLGSSVSAETEQQTSQVTKSEDDKA 60
QY 54 PSDPKFDA----QAQNGQLAQAMFKAANQADQTATSOVSPATDGRVDNQVTPAANQPAAN 109
Db 61 SESSQTDAPKTKQAQTEQ-TQAQ-SQANVAD-TSTS----ITKETPSQNIITQANSDDKT 113
QY 110 VANQVANPATDAGALNRQSAADTSTDGKAVP-----QTSDDPG-----H 149
Db 114 VNTKSEEAQTESEERTKQSEEAQTTASSQALTKQAKAELTKQRTAAQENKPNVDLAAIPN 173
QY 150 LETVDGKTYVDANGQRKLNKNSWIDGKTYYPDGOTGGAQTDLP-----KTGOANQDNVDP 205
Db 174 VQIDGKTYIYIGSDGQPKKNFALTNNKVLVPDKNTG-ALTDTSQYQFQGLTKLNN--- 229
QY 206 SYQANNOAYSNEASFETVDNYLTADSWYRPRKILKNGOSWASSEGDLRPLMTWWDPA 265
Db 230 DYTPEHQVINFENTSLTIDNYVTADSWYRPKDILKNGKWTASSEDRLPILMSWWDPK 289
QY 266 ATKAAYANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEKKIASEGNTNWLDRKMSQ 322
Db 290 QTQIAYLNMNQGLGTGENTYADSSQESLNLAQTVQVKIETKISQTOQTQWLARDINS 349
QY 323 FVKSONQHSIASNETVYPNODHMQGGLALLPSNSKDTSEHANSDWELLANRPTFTGKQKY 382
Db 350 FVKTPQNNWSQTESDTSAGEKDLHQGGALLYSNSDKTAYANSDYRLNRTPTSQTKPKY 409
QY 383 FTTNYA-GYELLANDVDNSNPVQAEQLNHLHYLMNWGDI VMGDKDANFDGVRVDAVDN 441
Db 410 PEDNSSGGYDFLLANDINDNSPQVQAEQLNWLHYLMNYGSI VANDPEANFDGVRVDAVDN 469
QY 442 VNADLLQTDYRYKAKYCTDQNEKNVADHLSILEAWSGNDNDYVQDNNFSLINDORS 501
Db 470 VNADLLQIASDYLKAHYGVKSEKNAI NLSILEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529
QY 502 GMLKAF-----GYASAYRGNLSNLATAGLKNRSA-NPDSDPVNPVYFIRAHSEVQT 552
Db 530 SLLYALTRPLEKQDASNKNEIRSGLEPVTITNSLNRSAGEKNSERMANVYFIRAHSEVQT 589
QY 553 RIAKIIRKLGKTNAGLNLTLDDLNKAFDIYNQDMNATDKVYPPNNLPMAYAWMLQNK 612
Db 590 VIAKIIKAQI-NPKTDGLT-FTLDELKQAFKIYNEQDMRQAKKKYQTSNIPTAYALMSNK 647
QY 613 DTVTRVYVYGDMDYDNGQYMATKTPFYNAETLLKGRIKYVAGGO--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYSDGQYMATKSPYDAIDTLKARIKAYAGGQDMKITYVGGDKSHMDW 707
QY 665 S-SGILTSVRYGKANSASDAGNTETNSGMALLNNRPNFRAYRN--LTLNNGAAHKSQ 721
Db 708 DYTGLVTSVRYGTGANEATDQSEATKTQGMVAVITSNPDLKQNDKVIYNMGAHKNQ 767
```

Qy	722	AYRPLLLSTKDGATATYLNDSGDVSRQYKYTDSQGNLSFSASELSQSVANAQVSGMIQWVP	761
Db	768	EYRPLLLTTXDGTSYTDAAAKS-LYRKNTDKGELVDFDASDIQQGLNPQVSYLAWVP	826
Qy	782	VGAADNQDVRTSPSTQATKGNLYHOSDALSQVIVEGFNFQAFAGSPDYTNVIAKN	841
Db	827	VGASDNQDVRVAASNKANATQGVYESSALDSQLIVEGFNFQDFTVKDSYTNKKIAQN	886
Qy	842	GDLFKSWGITQFEMAPQVYSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYGSKQDLANA	901
Db	887	VQLFKSWGVTSFEMAPQVYSSDGSPLDSIIQNGYAFEDRYDLAMSKNNKYGSSQDMINA	946
Qy	902	IKGLOSAGIKVLSDLVPNQLYNLPCKEVTWATRVNQYQOAKSATSINKTIPYVANTRSYG-	960
Db	947	VKALHKSIGIQVIAADWPDQIYNLPCKEVTWATRVNDYGEYRKDSEIKNTLYAAANTSNGK	1006
Qy	961	DYQEOYGCKFLDDLQKLYPLRFLSTKQISTCKPDPSPVKITNWSAKYFNGSNILLGRGAKYV	1020
Db	1007	DYQAKYGGAFLSLAAKYPISFNRTQISNGKKIDPSEKITAWKAKYFNGTNILLGRGVGYV	1066
Qy	1021	LSE--GNKYLNLDAGKULFLPTVLNNTYGGQPOVSANGPISKNGGIHYLDKNGQEVKORF-K	1077
Db	1067	LKDNASDKYPELKGNOTLYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ	1121
Qy	1078	EISGSWYYPDSGDMATGKTIKNDTYLFWPNGKQLKEGW--YDGGKAYYYDDNGRTWT	1135
Db	1122	DAGNWTYYFNNGHMVYGLQOLNGEVQYFUSNGVQJRESFLENADGSK-NYFCHLGNRYS	1180
Qy	1136	NKGFEVPRVDGQKWRYFNFGDGTIAIGLVLSDNRTLYFDAYGYQVKGQTVT-INGKSYP	1194
Db	1181	N-GYYSF--DNDSKWRYFDASGVMVAGLKTINGNTQYFDQDGYQVKGAWITGSDGKKRYF	1237
Qy	1195	DADGDLVQTDNANPAPQOGAGWKLLGDNQWGY-RKDQQLLTGEQITIDGQKVFQDNQVQ	1253
Db	1238	DDGSGNNAVRFANDK-----NGSDWYTLNSDGIALVGQTINGKTYTFYFGDGKQ	1286
Qy	1254	VKGTTATDASGLRIFYDRDQGHQVGKGWYSTSDDNWYVNESQVLTGLQITIDGQTVYFD	1313
Db	1287	IKGKIITD-NGKUKYFLANGSELARNIFATDSQNNWYFGSDGVAVTGSQTIAGKLYFA	1345
Qy	1314	DKGIQAKGKAVPMDENGNLRYFDADSGNMLDRWK-NYDGNWYFNRNGLA	1362
Db	1346	SDGKOVKGSFV-TYNGKVHYHADSGBLOVNRFEADKGNWYLDNSGEA	1394

RESULT 4

<hr/>	
US-08-793-824-2	
; Sequence 2, Application US/08793824	
; Patent No. 5981838	
; GENERAL INFORMATION:	
; APPLICANT: Simpson,	Christine Lynn
; APPLICANT: Giffard,	Philip Morrison
; APPLICANT: Jacques,	Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to	
; TITLE OF INVENTION: Increase Stored Carbohydrates	
; NUMBER OF SEQUENCES: 2	
; CORRESPONDENCE ADDRESS:	
; ADDRESSEE: Griffith Hack & Co	
; STREET: Level 8, 168 Walker Street	
; CITY: No. 5981838th Sydney	
; STATE: New South Wales	
; COUNTRY: Australia	
; ZIP: 2060	
; COMPUTER READABLE FORM:	
; MEDIUM TYPE: Floppy disk	
; COMPUTER: IBM PC compatible	
; OPERATING SYSTEM: PC-DOS/MS-DOS	
; SOFTWARE: PatentIn Release #1.0, Version #1.30	
; CURRENT APPLICATION DATA:	
; APPLICATION NUMBER: US/08/793,824	
; FILING DATE:	
; CLASSIFICATION: 800	
; PRIOR APPLICATION DATA:	

Qy	910	IKULSDLVPNQOLYNLPKEVWTATRVNQYQAKSGATINKTPYVANTRSYG-DYOEQYGG	968
Db	923	IKWMDWVPQMTAFPEKEVVVTRVDKGTVPVAGSQKNTLYVDGKSGKQQAQYGG	982
Qy	969	KFLDDLOKLYPRLPSTKQISTGKPIDPSVKITWNSAKYFNGSNILRGAKYVLSE--GNK	1026
Db	983	AFLEELQAKYPELFARQISTGVPMDSVKIKQWSAKYFNGTNILGRGAGVYLKQDATNT	1042
Qy	1027	YLNLAQK--LFPTVLNNTVGPQVBSANGFISKNQGIHYLDKNQVEQKVRKFISSGWY	1084
Db	1043	YFNISDNKEINFLPKTLN-----QDSQVGSYDQKGVVYSTSGYQAKNTFISEGDKWY	1097
Qy	1085	YFQSDGKMAKTKIGNDTYFLWPNGKQLEGVW--YDGKAYYYDDNGRTWTKGPFVEF	1142
Db	1098	YFDNNGFMVTAQSGINGVNYTFYLSNGQLKDAILKNEGDGVVAYYGNQ--GRYEN-GYIQF	1155
Qy	1143	RVGDQDKWRYFNGDGTITAIGLVSLDNRTLYFDAYGYQVKQTV--TINGKSYTFDADQGL	1201
Db	1156	M--SGVWRHFN-NGEMSVGLTVIDGVQVYFDEMGYQAKGFVTTADGKIRYFDPKQSGNM	1211
Qy	1202	VQTDNANPAQOQAG-WKLLGDNQWYRKDGQOLLTGEOTIDGKVFPQDNQYQVKGSTAT	1260
Db	1212	YR-----NRFIENEKGKLYLGE-----DGAAVTGSQTINGOHLYFRANGVQVKGFEVT	1260
Qy	1261	DASGLVLFYDRDQGHQVKGKMYSTSDNNVYVNESQVLTKLOITDQCTVYFDDKGIQAQK	1320
Db	1261	DHHGRISYDNGSDQJIRNFRVRNAQOMFYFDNNGYAVTGARTINGQLLYFRANGVQVK	1320
Qy	1321	GKAVMDENGNLRYFDADSGNMLDRW-KNVDGNWYYFNRNGLA	1362
Db	1321	GEFTYDRYGRISYDNGSGDQIRNFRVRNAQOMFYFDNNGYA	1363


```
QY 1321 GKAVDENGRLRYFDADSGMMLDRW-KNVVDGNWYFNRNGLA 1362
DB 1321 GBFTDTRYGRISYIDGNSGDQIRNRFVRNAQGWQFYFDNNGYA 1363

RESULT 8
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4
```

```
Query Match 40.0%; Score 2893.5; DB 2; Length 1375;
Best Local Similarity 44.3%; Pred. No. 4.9e-197;
Matches 634; Conservative 211; Mismatches 412; Indels 175; Gaps 30;
```

```
QY 1 MEKNLYKLHKVKOMVAGVTT--VTLSFLAGGVAAADTNNNGTSVQVKNWVPSDPK 58
DB 1 MEKKVRFKURKVKRWVTVSIAVVTLSLS-GSLVRADSTDDRQAVT-----49
QY 59 FPAQANGQOLAQAMFKAANOAD-OTATSQVSPATDGRVDN-QVTPAANOQPAANQDVA 116
DB 50 -ESQASLVTTSEAAKETLTATDTSTATSATSQPTATVTDNVSTTNGSTNTTANTAN-FVV 107
QY 117 NPA-----TDGALNRQSAADTSDGKAVPQTS-----145
DB 108 KPTTTSEQAKTDSKIIITTSKAVNRLTAT-----GKFVPANNNTAHPKVTVDKIVPIKP 162
QY 146 -----QPG-----HLETVDGKTVYVDANGORLKNYSWVIDGKTVYFDGQT 185
DB 163 KIGKUKQPSLSQDDIAALGNVKNRKVNGKTVYKEDGTLOKVALNGTKTFFDETG 222
QY 186 GEAOTDLP-KTGQANQVDPDSYQANNOQAYSEASSFETVDNYLTADSWYRPRKILKNGQ 244
DB 223 ALSNNLTLPKKGKNTNNDNTNSFAQNVQYSTVDANFEHVDHYLTAESWYRPKYILKDGK 282
QY 245 SHQASSEGDLRLPMTWTPDAATKAAAYANFWAKBGLISGSYQNSA--NLDAAATNIQSA 302
DB 283 TWTQSTEKDFRPLMTWTPDQETQRQYVNYMNAQLGIHQTYNTATSPQLNLAAQTITQK 342
QY 303 IEKKIASEGNTNWLKMSQFVKSONOWSIASENETVYPNQDHMQGALLFS-NSKDETH 361
DB 343 IEKITAENKTNWLRTISAFVKTQSAWNSDSEK-----PFDHLQKGALLYSNNKLSITQ 398
QY 362 ANSDWRLNLRNPTFTQTKO--KYFT--TNYAGYELLANDVDNSNPVQAEQLNHLHLYMN 418
DB 399 ANSNRYILNRTPTNQTGKDPRTYADRTTIGGYEFLFLLANDVDNSNPVQAEQLNHLHLYMN 458
```

RESULT 9

```
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
```

```
QY 419 WGDIVMGDKDANFDGVRVDVAVDNNVADLLQIORDVYKAKYGTDOEKNKAIDHLSILEAWS 478
DB 459 FGNIIYANDPDANFDSIRVDVADVNDADLLQAGDYLKAAKGLHKNDKAANDHLSILEAWS 518
QY 479 GNDNDYVKDQNNFSLSIDNDQSGMLKAFGASAYRGNLSNLSLATAGLKNRS-ANPDSDPV 537
DB 519 YNDTFYLHDDGDNMINDNRLRLSLLSLAKPLNQSGMNPPLITNSLVNRTDDNAETAAV 578
QY 538 PNYVFIRAHDSVQTRIAKIREKLGKTNADGLTNLTDLLNKAFDIYNOQMNADTKVY 597
DB 579 PSYSFIRAHDSVQDLIRNIIRTEI-NPNVVGY-SFTTEIKKAFEIYNKDLILATEKKYT 636
QY 598 PNNLEPMAWMLQNKDVTTRYVYGDMYTDNGQYMATKTPFYNAIETLKGRIKYVAGGQA 657
DB 637 HYNLTASYALLITNKSVPRTYVYDGMFTDDGGYMAHKTINYEAIETLLKARIKYVGGQA 696
QY 658 VSYKQDSSGILTSVRYGKGANSASDAGNTETNSGSMALLINNRRNFR--AYRNLTJLNMG 715
DB 697 MRNQOVGNSEIITSVRYGKGALKATDTCGTRTTSQVAVIEGNNLSLALKASDRVVNMG 756
QY 716 AAHKSQAYRPLLLSTKDGITATYLNDSVDVSQYKYTDSQNLSPSASELSQSVANAQVSGM 775
DB 757 AAHKNQAYRPLLLTNDNGIKAYHSDQEA-AGLVRYTNDRGELIFTAADIKGYANPQVSGY 815
QY 776 IQVWVPGAADNOVDRTSPSTQATKDGNIYHQSODALDSQVIYEGESNFQAFQASPDQYTN 835
DB 816 LGVWVPGAAAADQDVRVAASTAPSTDGKSVDHQAALDSRVMPFEGSNFOAFATKKEEYTN 875
QY 836 AVIAKNGDLFKSWGITQFEMAPQYVSSSEDTGFLDSVILNGVAFSDRYDLAMSNNKYGSK 895
DB 876 VVIACNVDKFAEWGVTDFEMAPQYVSSDTGFLDSVILNGVAFSDRYDLGISKPNKYGTA 935
QY 896 ODLANAIGLOSAGIKVLSLVNQLYNLPKEVVVATRVNOYQAKSGATINKTPYVAN 955
DB 936 DDLVKAIKALHSKGIKVMADWPDQMYALPEKEVVVATRVNRYKGPVAGSQIKNTLYYVD 995
QY 956 TRSYG-DYQEOYGGKFLDLQKLYPRLFTQISTGKPIDPSVKITNSAKYFNCSNIIIG 1014
DB 996 GKSSGKQQAQYGGAFLEELQAKYBELFARKQISTGVPMDSVKIKQWSAKYFNGNTNIIIG 1055
QY 1015 RGAKVYLSLSE--GNKYINLADGKFLFTVLNNTYQPVVSANGFISKNGGIHYLDKNGQBV 1072
DB 1056 RGAGVYLDKQATNTYFSLVSDNTFLPKSLVNPNHGCTSSSVTGLVFDGKGYVYTSNGQA 1115
QY 1073 KNRPKESGWSYVYFSDGKMATGKTIKNDTYLMPNKGQKLBGWWYDGKKAY-YDDNG 1131
DB 1116 KNAFISLGNWYVYFNNGYMVTGAQSIKNGANYFSLNGITQLRNASIYDNGNKVLSYVYDNG 1175
QY 1132 RTWTKGFVERVVDGQKWRYPNGDGTIAIGVSLDNRITLYPDAYGYQVKGQTVTINGKS 1191
DB 1176 RRYEN-GYILF---GQ-QWRYFQ-NGIMAVGLTRVHGAVQYVFDASGFOAKGQFI----- 1223
QY 1192 YTFDADQDGLVQTDNANPAQOQAGWKLIGDNQWYRKDGQLLGEQITDGOKVFPQDNG 1251
DB 1224 -----1223
QY 1252 VQVKGATDASGLVRFYDRDQGHQVKGWGTSTSDNNWYVNESQVLTGLTIGDQTVY 1311
DB 1224 -----TTADGKLYFDRDSGNQISNRFVRNSKGEWFLFDHNGVAVTGTVTFNGQRLY 1275
QY 1312 FDDKGIQAKGKAVMDENGLRYFDADSGNMLDRW-KNVVDGNWYFNRNGLA 1362
DB 1276 FKPNGVQAKGEFIRANGYLRYPDPSNGEVRNFRNRSKGEWFLFDHNGIA 1327
```



```
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3

Query Match      30.3%; Score 2191.5; DB 2; Length 1278;
Best Local Similarity 40.9%; Pred. No. 4.8e-147;
Matches 518; Conservative 170; Mismatches 436; Indels 141; Gaps 34;

QY 36 AADTNNNDGTSV--OVNK-MVPSDPKFDQAQNG--OLAQAMPKAAQADQATATSOVSPAT 91
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 AADGNSGNTVTDQFSKNYATTGNGFDYKVGNGQVFEFSGHATNQSDKDSQWIIIVL 137
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 92 DGR-VDNQVTPAANQPAANVANQDV--ANPATDAGAL-----NRQSAAD 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 NGKEVKRQLVNDTKEGAAGFNRNDVYKVPAINSSMSGFQGIITLPVTVKENVQLVHR 197
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 133 TSTDGKA-----VP-QTSDQPGH-----LETVDGKTYVD-ANGQRLKNYS 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 FNSDVKTGSGNYVDFWSELMPVKDSFQKNGPLKQFGLQTINGQQYYIDPTTGQPRKNFL 257
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 172 MVIDGKTYVFDGQTEAQTDLPKTQANQDNVPSYQANNOQAYSNEASSFETVDNYLTAD 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 258 LOSGNNWIIYFDSGTGVTGNALELOFAKGTVSNEQYRNGNAYSYDKSIENVNGYLTAD 317
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 232 SWYRPRKILKNGSQWQASSEGDLRPIILMTWPDATKAAYANFWAKEG-LISGSYRQNSA 290
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 318 TWYRPAQILKGTWTTDSKETDWRPILMVWMENTLTQAYLYNMKQHGNTLSALPFFNA 377
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 291 NIDATON-----IQSAIEKKIASEGNTWLRDKMSQFVKSONQWSIASNETVVPNQDH 345
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 378 DADPAELNHYSEIVQONIEKRISETGNTDRLTLAHDFVTNNPMWNKDSENVFSGIQ-- 435
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 346 MOGGALLFNSKDTSEANDHLLNRNPTFQTKQKFTTYNAGVELLALANDVDNSNPV 405
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 436 FOGGFLKYNSDLTPYANSDYLLGRMPI-----NIKQTYRGQFLLANDIDNSNPV 489
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 406 QAEQLNHLHYLNMWGDIVMGDKDANFDGVRDAVNANVADLLQIQDYKAKYGTQDNEK 465
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 490 QAEQLNWLIIYLLNFGTITANNQANFDSVRVADPNIDADLNIAQDYFNAAYGMD-SDA 548
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 466 NAIDLHLSLEAWGNDNVYQONNPSLIDNDQSGMLKAFGYASAYRGNLSNLTATAGL 525
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 549 VSNKHINILEDWNHADPEYFNKIGNPQLTMDDTIKNSLNHGLSDAT-NRWGLDAIVHQS 607
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 526 KNRSANPDSDPV-PNVVETRAHDSSEVOTIAKIREKLGKTNADGLTNLTDLNKAFDI 584
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 608 ADRENNSTENVVIPNYSFVRAHDNNSQDQIQAIRADVTK-----DYHTTFEDEQKGIDA 663
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 585 YNQDMNATDKVYYPNNLPMAWMLQNKDTVTRVYVYGDVMTDNGOVMATKTPFYNAIETL 644
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 664 YIQDQNSTVKKNLYNIPASYAILTNKDTIPRVYVYGLDYGQVMEHQTYRDTLNL 723
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 645 LKGRKYVAGQOAVSYKQDWSSG----ILTSVRYGKGANSASDAGNETRNSGMAILLNN 700
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 724 LKSRKYVAGGQSM--QTMVGGNNILTSVRYGKGAWTATDTGTDETRTQIGVYVSN 780
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 701 RNPFPAYRN--LTLMGAHKAQYRPLLLSTKDGIAITYLNSDSDVSQYKYTTDSQGNLS 758
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 781 TPNLKLGVNDKVVLHGAHKNQYRAAVLTTTGDVINYTSD---QGAPVAMTDENGDLY 837
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 759 FSASEL-----QSVANAQVSGMIQVWPVGAADNDVRTSGPSTQATKQGNLYH 806
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 838 LSSHNLVNGKEADTAQGYANPDVSGYLAWVPVPGASDNDQARTAFSTEKNSAYR 897
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 807 QSDALDSQVYBGFNFQAFQSPQYTNNAVIAKNGLDFKSWGITQFEMAPQYVSSSEDGT 866
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 898 TNAAFDNSVIREAFSNFVYPTTKESERANVRIAQADFFASLGFTSFEMAPQYNSSKORT 957
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 867 FLDSVILNGYAPSDRYDLAMSKNKYKSGQDLANAKGLQSGAGIKVLSLVPNQLYNLPG 926
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 958 FLDSIDNGYAPTDRYDLGMSFPNKYGTDEDLRNAIQALHKAGLQVMADWPDQIYNLPG 1017
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 12
US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDIINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAQUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-2

Query Match      30.3%; Score 2191.5; DB 2; Length 1781;
Best Local Similarity 40.9%; Pred. No. 8e-147;
Matches 518; Conservative 170; Mismatches 436; Indels 141; Gaps 34;

QY 36 AADTNNNDGTSV--OVNK-MVPSDPKFDQAQNG--OLAQAMPKAAQADQATATSOVSPAT 91
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 581 AADGNSGNTVTDQFSKNYATTGNGFDYKVGNGQVFEFSGHATNQSDKDSQWIIIVL 640
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 92 DGR-VDNQVTPAANQPAANVANQDV--ANPATDAGAL-----NRQSAAD 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 641 NGKEVKRQLVNDTKEGAAGFNRNDVYKVPAINSSMSGFQGIITLPVTVKENVQLVHR 700
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 133 TSTDGKA-----VP-QTSDQPGH-----LETVDGKTYVD-ANGQRLKNYS 171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 701 FNSDVKTGSGNYVDFWSELMPVKDSFQKNGPLKQFGLQTINGQQYYIDPTTGQPRKNFL 760
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 172 MVIDGKTYVFDGQTEAQTDLPKTQANQDNVPSYQANNOQAYSNEASSFETVDNYLTAD 231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 761 LOSGNNWIIYFDSGTGVTGNALELOFAKGTVSNEQYRNGNAYSYDKSIENVNGYLTAD 820
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 232 SWYRPRKILKNGSQWQASSEGDLRPIILMTWPDATKAAYANFWAKEG-LISGSYRQNSA 290
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 821 TWRPKQILKDGTTWDSKETDMRPILMWWNTLTQAYLLNMYKHGNLLPSALPPFNA 880
QY 291 NUDATQN-----IOSAIEKKIASGNTWLRKXMSQFVKSONQWSIASENETVVPNQDH 345
Db 881 DADPAELNHYSEIVQONIEKRISETGNTDLRLMHDFTVNNPMMNKDSENVFSGIQ-- 938
QY 346 MOGGALLFSNKSDEHANSDFLLNRPFTQKQKYPFTNYAGVELLNDVDSNPV 405
Db 939 FQGGFLKYENSULTPYANSDYLLGRMPI-----NIKQTYRGQEPFLANDIDNSPVV 992
QY 406 QAEQLNHLHYLNMWGDVGMGDNDANFDGVRDADVNADLNLIQIQORDYKAKYGTQNEK 465
Db 993 QAEQLNWLYYLNFGTITANNQANFDSVRDAPDNIDADLNIAQDYENAYAGND-SDA 1051
QY 466 NAIDLHLSLEAWSGNDVYKQDNFSLIDNDQSRGMLKAFYASAYRGNLSNLTATGL 525
Db 1052 VSNKHINILEDWNHDEYFNKIGNPQLTMDDTIKNSLNHLGLSDAT-NRWGLDAIVHQS 1110
QY 526 KRSANPDSDPV-PNVVFTIRAHDSVOTRIAKIIEKLGKTWADGLTNLTDLNKAQDI 584
Db 1111 ADRENNSTENVIPNYSFVRAHDNNSQDQIQAIRHDVTGK----DYHTTFDEQKGIDA 1166
QY 585 YNQDMNATDKVYPNNLPMAYAWMLQNDTVTRVYGDVMTONGQVMAKTFPFYNAIETL 644
Db 1167 YIQDNSTVKKNLNIPASYAILTNKDTIPRVYGDLYTDGGQYMEHQTRYDTLNL 1226
QY 645 LKGRIKYAGGQVAYKQDWSG-----ILTSVRYGKGANSASDAGNETRNSGMLINN 700
Db 1227 LKSRVKYAGGQSM---QTMVGNNILTSVRYGKGAMTATDTGTDTRTQGIQVWVS 1283
QY 701 RNFPRAYRN--LTLNNGAAKQAYRPLLSLTKDGIATYLNDSVDVSRQKYKTDQGNLS 758
Db 1284 TPNLKLGVNDKVVLHMGAAHKNQYRAAVLTITTDGVINTSD---QGAPVAMTDENGDI 1340
QY 759 FSASEL-----QSVANAQVSGMIQVWVPVGAADNODVRTSPSTQATKQGNVYH 806
Db 1341 LSHNLVNGKEADTAVQYANPDVSGVLAWVPVGAASNQDARTAPSTEKNSGNSAYR 1400
QY 807 QSDALDSQVIYEGFNFQAFQSPDQYTNVIAKNGDLFKSWGITQFEMAPQYVSEBGT 866
Db 1401 TNAAFDSNVIFEAFSNFVYPTPKESERANVRIAQNADFPASLGFTSFEMAPQYNSKORT 1460
QY 867 FLDSVLINGYAFSDRYDLAMSKNNKYKQDLANAIKGLQSAKIGVLSLDPNOLYNLPG 926
Db 1461 FLDSITDNGYAFTRDYDLGMSBPNKYGTDEDLRNAIQALHKLQAGLQVMAWVDPQIYNLPG 1520
QY 927 KEVVTATRVNQYQAKSGATINKTPVANTRSYGDYQEQYGGKFLDLDLQKLYPRLFSTKQ 986
Db 1521 KEVATVTRVDRGNVWKDAIINNLYVNTIIGGEYQKYGGAFLDKLQLYPEFTKKQ 1580
QY 987 ISTGKPIDPSKITNWSAKYFNGSNILGRGAKYVL--SEGNKYLN--ADGKLFPTVLNNT 1044
Db 1581 VSTGVAIDPSQKITEWSAKYFNGTNILHRGSGVYLVKADGGQYVNLGTTTKQFLPIQLT-- 1638
QY 1045 YQCPQVSANGFTSKNGIHYL--DKNGQEVKNRFEKIS--CSWYTFPSDGMATGKTKIGND 1102
Db 1639 -GEKQGNBGFVKGNDGNYFYFDLAGNMVKNFTFIEDSVGNWYFFDQDGKRVNKHFDVVD 1697
QY 1103 TYLFPNGKQKGVWYDGKK--AYYDDNGRWTNKGFFVFRVGDGKWRVFNPGDGTIAI 1161
Db 1698 SY-----GEKGYTFPLKNG-----VSFRG----- 1716
QY 1162 GLVSLDNRTLYFDAYQYKVGQTVTINGKSYTFDADQGDVLQVTDNANPA--FQGAQWKL 1219
Db 1717 GLVQTDNGTYFYDNYGKVRNQTINAGAMIYTLN-ENGLIKASYNSDAEYPTSTDVGKM 1775
QY 1220 LGDNQ 1224
Db 1776 LDQNK 1780
```

RESULT 13

US-09-604-957-4

```
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAULI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-604-957-4
```

Query Match 22.3%; Score 1615.5; DB 2; Length 545;
Best Local Similarity 58.9%; Pred. No. 1.4e-106;
Matches 323; Conservative 75; Mismatches 127; Indels 23; Gaps 9;

```
QY 393 LLANDVNSNPVQAEQLNHLHYLNMWGDVGMGDNDANFDGVRDADVNADLNLIQIQRD 452
Db 1 LLANDIDNSNPVQAEQLNHLHYLNMWGDVGMGDNDANFDGVRDADVNADLNLIQIASD 60
QY 453 YVKAQYGTQDQNEKNAIDHLSTLEAWSGNDNDVYKQDNFSLIDNDQSRGMLKAF----- 507
Db 61 YLKAHYGVDSKSEKNAIHLSTLEAWSDNDPNYKDTKGAQLPIDNKLRLSLYALTRPLE 120
QY 508 ---GYASAYRGNLSNLTATAGLKNRSA--NPDSDPVNPVYFIRAHDSVOTRIAKIIEK 563
Db 121 KDSANKNEIRSGLEPVITNSLNRSAGKNSERMANYIFIRAHDSVOTVIAKIIKAQI- 179
QY 564 KTNADGLTNLTDLNKAQFIDYQDNMATDKVYPNNLPMAYAWMLQNDKOTVTRVYGD 623
Db 180 NPKTDGLT-FTLDELKQAFKIYNEQDMRQAKKQYTSQNTPTAYALMLSNKSITRLYGD 238
QY 624 YTDNGQYMATKTPFYNAJETLLKGRIKYVAGQ--AVSYKQ-----DWS-SGILTSVRY 674
Db 239 YSDDQYMATKSPYDAIDTLKARIKAAAGQDMKIYVEGDKSHMDWDYTGVLTSVRY 298
QY 675 GKANSASDAGNETRNSGMLINNRPFRAYRN--LTLNNGAAHKSQAYRPLLLSTKD 732
Db 299 GTGANEATDQGEATKTCQMAVITSNPNFSLKLNQNDKVIIVNMGAAHKNQYRPLLLTTKD 358
QY 733 GIATYLNDSVDVSRQKYKTDQGNLSFSASELQSVANAQVSGMIQVWVPVGAADNODVRT 792
Db 359 GLTSYTSDAAKS--LYRKTNDKGLVFDASDIQGLNPNQVSGVLAWVPVGAASDQDVRV 417
QY 793 SPSTQATKQGNLYHOSDALDSQVIYEGFNFQAFQSPDQYTNVIAKNGDLFKSWGITQ 852
Db 418 AASNKANATQGVYESSSALDSQIYIEGFSNFDQFTKQSDYTNKIAQNVQLFKSWGITS 477
QY 853 FEMAPQYVSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYKQDLANAIKGLQSAKIGV 912
Db 478 FEMAPQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYKQDQMINAVKALHKSIGV 537
QY 913 LSDLVNQ 920
Db 538 IADWVPDQ 545
```

RESULT 14

US-09-995-749A-10

; Sequence 10, Application US/09995749A

; Patent No. 6867026

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-995-749A-10

Query Match 22.1%; Score 1597.5; DB 2; Length 545;
Best Local Similarity 58.8%; Pred. No. 2.7e-105;
Matches 323; Conservative 76; Mismatches 125; Indels 25; Gaps 11;
QY 393 LLANDVNSNPVVOAQLNHLHLMWGDIVMGDKDANFDGVRVDAVDVNNADLLQIQRD 452
Db 1 LLANDVNSNPVVOAQLNHLHLMWGDIVMGDKDANFDGVRVDAVDVNNADLLQIQASD 60
QY 453 YKAKYGTONEKNAIDHLSILEAWGNDNDYVKDQNNFSLSDNDORSGLKAFV 507
Db 61 YLKAHYGVDSKSEKNAIHLSILEAWGNDNDYVKDQNNFSLSDNDORSGLKAFV 120
QY 508 ---GYASAVRGNLSNATAGLKNRSA-NPDSDPVPNVYFIRAHDSVQTRIAREKLG 563
Db 121 KQASNKNEIRSGLEPVITNSLNRSAGKNSRMANYIFIRAHDSVQTRIAREKLG 179
QY 564 KTNADGLTLLDNLKAFDIYNQDNATDKVYPNNLPYAWMLQNKDTVTRVYVYGD 623
Db 180 NPKTDLGT-FTLDELKQAFKIYNEDMRQAKKYTQSNIPYAYLMLSNKDSITRLYYGDM 238
QY 624 YTDNGQYMATKPFYNAIETLLKRIKYVAGGO--AVSYKQ-----DWS-SGILTSVRY 674
Db 239 YSDGQYMATKPFYDAITLLKARIKYAAGGQDKITVEGDKSHMDNDYTGVLTSVRY 298
QY 675 GKGANASDAGNTETNSGALLINRPNFRYRN--LTLNMGAAHKSQAYRPLLLSTKD 732
Db 299 GTGANEATQGSSEATKTQGMVITSNNPSLKNQNDKVIWNMGAAHKNQYRPLLLTTKD 358
QY 733 GIATYLNDSVDVSRQKYTDSQGNLSFSASELQSV-ANAQVSGMIQVWVPVGAADNQDVR 791
Db 359 GLTSYTSDDAAKS-LYRKTNDRKGELVFDASDIQGLYLNLPQVSG-LAVWVPVGAADNQDVR 416
QY 792 TSPSTQATKQGNINYHOSDALDSQVIYEGFSNFQAPQSPQVYTNVIAKNGDLFKSWGIT 851
Db 417 VAASNKANATQVYSSSSALDSQLIYEGFSNFQDFVTKDSYDYNKKIAQNVLFSWGV 476
QY 852 QFEMAPQVYSSDGTFLDSVILNGYAFSDRYDLAMSNNKKNYKSKQDLANAIKGLQASG 911
Db 477 SFEMAPQVYSSDGSGLDLSIIQNGYAFEDRYDLAMSNNKKNYKSGQDMLNVAKHLKSG 536
QY 912 VLSDLVPNQ 920
Db 537 VIADWVPDQ 545

RESULT 15
US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN

; FILE REFERENCE: B0 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-5

Query Match 18.8%; Score 1362.5; DB 2; Length 523;
Best Local Similarity 51.8%; Pred. No. 1.5e-88;
Matches 275; Conservative 87; Mismatches 158; Indels 11; Gaps 6;
QY 393 LLANDVNSNPVVOAQLNHLHLMWGDIVMGDKDANFDGVRVDAVDVNNADLLQIQRD 452
Db 1 LLANDVNSNPVVOAQLNHLHLMWGDIVMGDKDANFDGVRVDAVDVNNADLLQIQAD 60
QY 453 YKAKYGTONEKNAIDHLSILEAWGNDNDYVKDQNNFSLSDNDORSGLKAFV 512
Db 61 YKFLAYGVDPNDATANQHLNLSILEWGHNDPLYYTDDQSNQLTMDYYVHTQLIWSLTKSSD 120
QY 513 YRGNLSNATAGLKNRS-ANPDSDPVPNVYFIRAHDSVQTRIAREKLGKTNADGLT 571
Db 121 IKGTMQRFVDYVWDRSNDSTENEALPNYSFVRAHDSVQTVIAQIVSDY--PDVNSL 178
QY 572 NTLDDLKNAFDIYNQDNATDKVYPNNLPYAWMLQNKDTVTRVYVYGDYTDNGQYM 631
Db 179 APTTEQLAAAFKYNEDEKLADKKYQYNMASAYAMLLTNKDTVPRVYVGLYTDGQYM 238
QY 632 ATKTPFYNAIETLLKRIKYVAGGOAVSYKQDWSGILTSVRYKGANASDAGNTETRN 691
Db 239 ATKSPYDAINTLLKARQVYVAGGQSMVD---SNDVLTSVRYGKQAMTASDTGTSETRT 295
QY 692 SGWALLINRPNFRAY--RNLTLNMGAAHKSQAYRPLLLSTKGIATYLNDSVDVSRQYK 749
Db 296 EGIGVIVSNNAELQLEDGHTVTLHMGAAHKNQAYRALLSTTADGLAY--DTD-ENAPVA 352
QY 750 YTDSQGNLSFSASELQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKQGNINYHOSD 809
Db 353 YTDANGDLIFTNESIVGVQNPQVSGYLAWVPVGAQDQDARTASDTTNTTSKVFHSA 412
QY 810 ALDSQVIYEGFSNFQAPQSPQVYTNVIAKNGDLFKSWGITQFEMAPQVYSSDGTFLD 859
Db 413 ALDSQVIYEGFSNFQAFATDSSEYTNVIAQNAQDFKQNGVTSFQLAPQVRSSTDTSF 472
QY 870 SVILNGYAFSDRYDLAMSNNKKNYKSKQDLANAIKGLQASGIVKLSLVPNO 920
Db 473 SIIQNGYAFSDRYDLGYGTPTKYGTADQURDAIKALHAGSIQAIADWVPDQ 523

Search completed: February 11, 2006, 19:46:10
Job time : 37.8957 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 122.176 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-39

Perfect score: 7230

Sequence: 1 MEKNLYKLVKVKQWVAIG.....WKNVDGNWYFNRNGLATRW 1365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7230	100.0	1365	4	US-10-383-930-39
2	7230	100.0	1365	5	US-10-797-821-39
3	3174	43.9	1430	3	US-09-740-274-6
4	3174	43.9	1430	4	US-10-383-930-36
5	3174	43.9	1430	5	US-10-797-821-36
6	3080.5	42.6	1554	4	US-10-383-930-38
7	3080.5	42.6	1554	5	US-10-797-821-38
8	3031	41.9	1590	4	US-10-383-930-37
9	3031	41.9	1590	5	US-10-797-821-37
10	2945.5	40.7	1475	3	US-09-740-274-2
11	2945.5	40.7	1475	4	US-10-383-930-34
12	2945.5	40.7	1475	5	US-10-797-821-34
13	2893.5	40.0	1375	3	US-09-740-274-4
14	2893.5	40.0	1375	4	US-10-383-930-35
15	2893.5	40.0	1375	5	US-10-797-821-35
16	2804	38.8	1518	4	US-10-383-930-40
17	2804	38.8	1518	5	US-10-797-821-40
18	2680.5	37.1	1497	5	US-10-484-218-18
19	2399	33.2	1595	5	US-10-484-218-20
20	2247.5	31.1	1777	5	US-10-484-218-12
21	2227	30.8	2057	4	US-10-417-280A-2
22	2191.5	30.3	1781	3	US-09-995-749A-2
23	2142.5	29.6	1771	5	US-10-484-218-14
24	2104	29.1	1006	5	US-10-484-218-22
25	1597.5	22.1	545	3	US-09-995-749A-10
26	1361	18.8	522	3	US-09-995-749A-11
27	1203.5	16.6	535	3	US-09-995-749A-13

RESULT 1

US-10-383-930-39
; Sequence 39, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 39
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-383-930-39

Query Match	100.0%	Score	7230;	DB	4;	Length	1365;
Best Local Similarity	100.0%	Pred. No.	0;				
Matches	1365;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MEKNLYKLVKVKQWVAIGVTTVTL	SLAGGVAAADTNNNDGTSVQVKNVPSDPKFD	60			
Db	1	MEKNLYKLVKVKQWVAIGVTTVTL	SLAGGVAAADTNNNDGTSVQVKNVPSDPKFD	60			
Qy	61	AAQNGQLAQAAMFKAANOADOTATS	QVSPATDGRVDNQVTPAANQPAANQVDPAT	120			
Db	61	AAQNGQLAQAAMFKAANOADOTATS	QVSPATDGRVDNQVTPAANQPAANQVDPAT	120			
Qy	121	DAGALNRQSAADTSTDGKAVPQTS	DQPHLETVDGKTYVDANGORLKNYSVMDGKTY	180			
Db	121	DAGALNRQSAADTSTDGKAVPQTS	DQPHLETVDGKTYVDANGORLKNYSVMDGKTY	180			
Qy	181	FDGQTGAQTDLPTKGANQDNVPS	YOANNOAVSNEASSPETYDNYLTADSWYPRKIL	240			
Db	181	FDGQTGAQTDLPTKGANQDNVPS	YOANNOAVSNEASSPETYDNYLTADSWYPRKIL	240			
Qy	241	KNGQSWQASSSGDLRPLMTW	PDAAATKAAAYANFWAKEGLISGSYRQNSANLDAATONIQ	300			
Db	241	KNGQSWQASSSGDLRPLMTW	PDAAATKAAAYANFWAKEGLISGSYRQNSANLDAATONIQ	300			
Qy	301	SAIEKKIASEGNTNWLDRKMSQ	FVKSQNQWSIASNETVYPNQDHMQGALLFNSKQDTE	360			
Db	301	SAIEKKIASEGNTNWLDRKMSQ	FVKSQNQWSIASNETVYPNQDHMQGALLFNSKQDTE	360			

ALIGNMENTS

28	1175.5	16.3	787	5	US-10-484-218-16	Sequence 16, Appl
29	1171	16.2	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1115.5	15.4	525	5	US-10-484-218-23	Sequence 23, Appl
31	554.5	7.7	223	5	US-10-484-218-6	Sequence 6, Appl
32	540	7.5	224	5	US-10-484-218-4	Sequence 4, Appl
33	532.5	7.4	223	5	US-10-484-218-10	Sequence 10, Appl
34	480.5	6.6	221	5	US-10-484-218-8	Sequence 8, Appl
35	469.5	6.5	221	5	US-10-484-218-2	Sequence 2, Appl
36	365.5	5.1	2710	4	US-10-011-366-6	Sequence 6, Appl
37	365.5	5.1	2710	4	US-10-354-774-6	Sequence 6, Appl
38	365.5	5.1	2710	4	US-10-271-012-6	Sequence 6, Appl
39	365.5	5.1	2710	4	US-10-729-122-6	Sequence 6, Appl
40	365.5	5.1	2710	4	US-10-729-039-6	Sequence 6, Appl
41	365.5	5.1	2710	5	US-10-729-527-6	Sequence 6, Appl
42	365.5	5.1	2710	5	US-10-727-898-6	Sequence 6, Appl
43	365.5	5.1	2710	6	US-10-728-696-6	Sequence 6, Appl
44	365.5	5.1	2710	6	US-11-001-241-6	Sequence 6, Appl
45	327.5	4.5	2366	4	US-10-011-366-10	Sequence 10, Appl

```
QY 361 HANSWRLNLRNPTQTQKQKFTTNYAGYELLANDVNSNPVQAQLNHLHYLMNWG 420
Db 361 HANSWRLNLRNPTQTQKQKFTTNYAGYELLANDVNSNPVQAQLNHLHYLMNWG 420
QY 421 DIVMGDKANFDGVRDADVNNADLLQIQRDYKAKYGTQDNEKNAIDHLSILEAWSGN 480
Db 421 DIVMGDKANFDGVRDADVNNADLLQIQRDYKAKYGTQDNEKNAIDHLSILEAWSGN 480
QY 481 DNDYVKQDNFSLSDNDQSRGMLKAFGYSAYRGNLSNLTAGLKNRSANPDSPPVNY 540
Db 481 DNDYVKQDNFSLSDNDQSRGMLKAFGYSAYRGNLSNLTAGLKNRSANPDSPPVNY 540
QY 541 VFIRAHSEVQTRIAKIIREKLGKTNADGLTNLTLDLNLKAPDIYNQDMNATDKVYPPN 600
Db 541 VFIRAHSEVQTRIAKIIREKLGKTNADGLTNLTLDLNLKAPDIYNQDMNATDKVYPPN 600
QY 601 LPMAYAWMLQNKDVTTRYVYGDYTDNGOYMATKTPFYNAIETLLKGRIKYVAGQAVSY 660
Db 601 LPMAYAWMLQNKDVTTRYVYGDYTDNGOYMATKTPFYNAIETLLKGRIKYVAGQAVSY 660
QY 661 KQDWSGILTSVRYGKGANSASDAGNTETRNSGMALLNNRPNFRAYRNLTLNMGAAHS 720
Db 661 KQDWSGILTSVRYGKGANSASDAGNTETRNSGMALLNNRPNFRAYRNLTLNMGAAHS 720
QY 721 QAYRPLLLSTKDGIAITYLNSDSDVSRQYKYTDSQNLSPFSASELOSANAVQVSGMIQVW 780
Db 721 QAYRPLLLSTKDGIAITYLNSDSDVSRQYKYTDSQNLSPFSASELOSANAVQVSGMIQVW 780
QY 781 PVGAADNDVRTSPSTQATKGNIIYHQSADLSDQVYEGFNSQFAFAPSQOYTNAVIAK 840
Db 781 PVGAADNDVRTSPSTQATKGNIIYHQSADLSDQVYEGFNSQFAFAPSQOYTNAVIAK 840
QY 841 NGDLFPKSGIITOFEMAPQVVSSEDTFLDSVLINGAFSDRYDLAMSNNKYGSKQDLAN 900
Db 841 NGDLFPKSGIITOFEMAPQVVSSEDTFLDSVLINGAFSDRYDLAMSNNKYGSKQDLAN 900
QY 901 AIKGLQSAIKVLSLDPNQLYNLPKGEVVTATRVNQYQAKSGATINKTPVANTRSYG 960
Db 901 AIKGLQSAIKVLSLDPNQLYNLPKGEVVTATRVNQYQAKSGATINKTPVANTRSYG 960
QY 961 DYQEOYGGKFLDLDLQKLPRLPSTQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKV 1020
Db 961 DYQEOYGGKFLDLDLQKLPRLPSTQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKV 1020
QY 1021 LSEGNKYLNLADGKFLPTVLNNTYQPOVSANGFISKNGGIHYLDKNGQEVKNRFXEIS 1080
Db 1021 LSEGNKYLNLADGKFLPTVLNNTYQPOVSANGFISKNGGIHYLDKNGQEVKNRFXEIS 1080
QY 1081 GSWYTFDSGKMATGKTKIGNDTYLFMPNGKOLKEGVMYDGGKAYYYDDNGRTWTKGFV 1140
Db 1081 GSWYTFDSGKMATGKTKIGNDTYLFMPNGKOLKEGVMYDGGKAYYYDDNGRTWTKGFV 1140
QY 1141 EFRVDGQKWRVFNCGDGTIAIGLVLSDNRTLYFDAYGVQVKGQVTVINGKSTFFDADQGD 1200
Db 1141 EFRVDGQKWRVFNCGDGTIAIGLVLSDNRTLYFDAYGVQVKGQVTVINGKSTFFDADQGD 1200
QY 1201 LVQTDNANPAPOQAGWKLKLDGNGYRDKDGLLTGEQTIDGQVFFQDNGVQVKGAT 1260
Db 1201 LVQTDNANPAPOQAGWKLKLDGNGYRDKDGLLTGEQTIDGQVFFQDNGVQVKGAT 1260
QY 1261 DASGVLRFVDRDQGHQVKGWYSTSDNNWVYVNESQVLTGLQTTIDGQTVFPDDKGIQAK 1320
Db 1261 DASGVLRFVDRDQGHQVKGWYSTSDNNWVYVNESQVLTGLQTTIDGQTVFPDDKGIQAK 1320
QY 1321 GKAVDENGRLRYFPADSGNMLRDRKKNVDGNNWYFNRNGLATRW 1365
Db 1321 GKAVDENGRLRYFPADSGNMLRDRKKNVDGNNWYFNRNGLATRW 1365
```

RESULT 2

US-10-797-821-39

; Sequence 39, Application US/10797821

```
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-797-821-39
```

```
Query Match 100.0%; Score 7230; DB 5; Length 1365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKNLYKHLHKQWVAIGVTTVTLSPLAGQVVAADTNNNDGTSVQVKNVPSDPKPD 60
Db 1 MEKNLYKHLHKQWVAIGVTTVTLSPLAGQVVAADTNNNDGTSVQVKNVPSDPKPD 60
QY 61 AOQNGQLAQAMFKAANOADQATATSOVSPATDGRVDNQTTPAANOPAANVANOVDANPAT 120
Db 61 AOQNGQLAQAMFKAANOADQATATSOVSPATDGRVDNQTTPAANOPAANVANOVDANPAT 120
QY 121 DAGALNRQSAADTSDGKAVPQTSQDPGHLETVDGKTYVDANGORLKNYSWVIDGKTY 180
Db 121 DAGALNRQSAADTSDGKAVPQTSQDPGHLETVDGKTYVDANGORLKNYSWVIDGKTY 180
QY 181 FDGQTEAQTDLPKTGQANQNDVPDSYOANNOAYSNEASSFETVDNYLTADSWYPRKIL 240
Db 181 FDGQTEAQTDLPKTGQANQNDVPDSYOANNOAYSNEASSFETVDNYLTADSWYPRKIL 240
QY 241 KNGOSWQASSEGDLRPIILMTWMPDAATKAAAYANFWAKEGLISGSYRONSANLDAATQIQ 300
Db 241 KNGOSWQASSEGDLRPIILMTWMPDAATKAAAYANFWAKEGLISGSYRONSANLDAATQIQ 300
QY 301 SAIEKKIASEGNTWLRDKMSQFVKSQNWSIASENETVYPNQDHMQGALLFSNSKDTE 360
Db 301 SAIEKKIASEGNTWLRDKMSQFVKSQNWSIASENETVYPNQDHMQGALLFSNSKDTE 360
QY 361 HANSWRLNLRNPTQTQKQKFTTNYAGYELLANDVNSNPVQAQLNHLHYLMNWG 420
Db 361 HANSWRLNLRNPTQTQKQKFTTNYAGYELLANDVNSNPVQAQLNHLHYLMNWG 420
QY 421 DIVMGDKANFDGVRDADVNNADLLQIQRDYKAKYGTQDNEKNAIDHLSILEAWSGN 480
Db 421 DIVMGDKANFDGVRDADVNNADLLQIQRDYKAKYGTQDNEKNAIDHLSILEAWSGN 480
QY 481 DNDYVKQDNFSLSDNDQSRGMLKAFGYSAYRGNLSNLTAGLKNRSANPDSPPVNY 540
Db 481 DNDYVKQDNFSLSDNDQSRGMLKAFGYSAYRGNLSNLTAGLKNRSANPDSPPVNY 540
QY 541 VFIRAHSEVQTRIAKIIREKLGKTNADGLTNLTLDLNLKAPDIYNQDMNATDKVYPPN 600
Db 541 VFIRAHSEVQTRIAKIIREKLGKTNADGLTNLTLDLNLKAPDIYNQDMNATDKVYPPN 600
QY 601 LPMAYAWMLQNKDVTTRYVYGDYTDNGOYMATKTPFYNAIETLLKGRIKYVAGQAVSY 660
```

Db 601 LPMAYAWMLONKDTVRVYIGDMYTNGQTMATKTPFYNAIETLLKGRKYYVAGQAVSY 660
Qy 661 KQDSSGILTSVRYGKANSASDAGTETRNQSMALLINNRPNFRAYRNLTLNMGAAHKS 720
Db 661 KQDSSGILTSVRYGKANSASDAGTETRNQSMALLINNRPNFRAYRNLTLNMGAAHKS 720
Qy 721 QAYRPLLLSTKQGIATYLNDSVDSDQYKYTDSQGNLSFASASELQSVANAQVSGMIQVWV 780
Db 721 QAYRPLLLSTKQGIATYLNDSVDSDQYKYTDSQGNLSFASASELQSVANAQVSGMIQVWV 780
Qy 781 PVGAABNDQVRTPSPQATKQGNHYHOSDALDSQVIEGFSNFOAFAQSPDOVTNAVIK 840
Db 781 PVGAABNDQVRTPSPQATKQGNHYHOSDALDSQVIEGFSNFOAFAQSPDOVTNAVIK 840
Qy 841 NGDLFKSWGITQFEMAPQVSSSEDTFLDSVILNGYAFSDRYDRLAMSKNKYGSQDLAN 900
Db 841 NGDLFKSWGITQFEMAPQVSSSEDTFLDSVILNGYAFSDRYDRLAMSKNKYGSQDLAN 900
Qy 901 AIKGLQASIKVLSLQNLNLPKQEVVTTATRVNQYQOAKSGATINTKPYVANTRSYG 960
Db 901 AIKGLQASIKVLSLQNLNLPKQEVVTTATRVNQYQOAKSGATINTKPYVANTRSYG 960
Qy 961 DYQEOYGGKFLDQKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 961 DYQEOYGGKFLDQKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Qy 1021 LSEGNKYNLADKGLFLPVLNNTYQPOVSANGFISKNGGIHYLDKNGQEVNRFKELS 1080
Db 1021 LSEGNKYNLADKGLFLPVLNNTYQPOVSANGFISKNGGIHYLDKNGQEVNRFKELS 1080
Qy 1081 GSWYFDSGKMATKTKIGNDTYLFMPNGKQKQGVYDGGKAYYDDNGRTWTKGPFV 1140
Db 1081 GSWYFDSGKMATKTKIGNDTYLFMPNGKQKQGVYDGGKAYYDDNGRTWTKGPFV 1140
Qy 1141 EFRVDGQKWRVFNQDGTITAGLVSIDNRTLYFDAYGYQVKQOTVINGKSYTFDADQGD 1200
Db 1141 EFRVDGQKWRVFNQDGTITAGLVSIDNRTLYFDAYGYQVKQOTVINGKSYTFDADQGD 1200
Qy 1201 LVQTDNANPAPQOAGWKLGNQNGYRKDGQLLTGEQITDGQVFFQDNGVQVKGGTAT 1260
Db 1201 LVQTDNANPAPQOAGWKLGNQNGYRKDGQLLTGEQITDGQVFFQDNGVQVKGGTAT 1260
Qy 1261 DASGVLRYDRDQGHQVKGWYSTSDNNWYVNESGQVLTGLQITIDGQVTFDDKGIQAK 1320
Db 1261 DASGVLRYDRDQGHQVKGWYSTSDNNWYVNESGQVLTGLQITIDGQVTFDDKGIQAK 1320
Qy 1321 GKAVNDENGLRYFDADSGNMLRDRWQVNDGNWYFNRNGLATRW 1365
Db 1321 GKAVNDENGLRYFDADSGNMLRDRWQVNDGNWYFNRNGLATRW 1365

RESULT 3
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. US2002031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172

; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match 43.9%; Score 3174; DB 3; Length 1430;
Best Local Similarity 47.3%; Pred. No. 1.7e-182;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;
Qy 1 MEKNRYKLKVKQWVAIGVTT--VTLFLAGGQVVAADTNNNDGTSV-----QVNMV 53
Db 1 METKRYKMKVKWVTVAVASGLITITGLTGLSSVSAETEQTSDVKVTKSEDDKAA 60
Qy 54 PSDPKFDA---QAQNGQLAQMFKAAQADOTATISQVSPATDGRVDNOVTPAANQPAAN 109
Db 61 SESSQTDAPKTKQAQTEQ--TQAQ--SQANVAD--TSTS-----ITKETPSQNIITTOANSDDKT 113
Qy 110 VANQDVANPATDAGALNROSADTSTDGKAVP-----QTSQPG-----H 149
Db 114 VTNKSEEAQTSEERTKQSEEAQTASSQALTKQAKAELTKQRTAAQENKNPVDLAAIPN 173
Qy 150 LETVPGKTYVVDANGRLKNYSWVIDGKTYYPDGGTGEAQTDLP-----KTQGANQDNVDP 205
Db 174 VQIIOGKYYIIGSDQPKPKNFALTNNKVLVYFDKNTG--ALTDTSQVQFKGGLTKLAN-- 229
Qy 206 SYQANNOAYSNEASSFETVDNYLTADSWYRPRKILKNGSQWASSEGLRPIILMTWMPDA 265
Db 230 DYTPEHQVNFENTSELTIDNVTADSWYRPRKILKNGKTWTATSESDLRPLMLSMWPK 289
Qy 266 ATKAAVANFAWEGLISG--SYRQNSA--NLDAATQNIQSAIEKKIASEGNTNWLKMSQ 322
Db 290 QTQIAYLNMNQOGLGTGENTYADSSQESLNLAQTVQVKIETKISQTOQTQWLRLDIINS 349
Qy 323 FVKSONQNSIASENETVYPNDDHMOGGALLFSNKSDEHANSDRLLNRPITFOTCKQY 382
Db 350 FVKTPNNNSQTESDTSAGEKDLQGGALLYSNKTAYANSYDRLLNRPITFOTCKPKY 409
Qy 383 FTTNYA-GYELLANDVNSNPVQAEQLNHLHYLMNMGDIVMGDKDANFDGVRVDAVN 441
Db 410 FEDNSSGGVDFLLANDIDNSNPVQAEQLNHLHYLMNYGSIIVANDEANFDGVRVDAVN 469
Qy 442 VNADILLQTDYKAKYGTQDQNEKVAIDHLSILEAWSGNDNDYVKDQNNFSLSDNDQRS 501
Db 470 VNADILLQIASDYLKAHYGVDKSEKNAIHNLSILEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529
Qy 502 GMLKAF-----GYASAYRGNLSNLATAGLKNESA-NPDSDPVPNTYVFIABHSEVOT 552
Db 530 SLLYALTPLPEKADASKNEIRSGLEPFVTNLSNRSABGKNSERMANYIFIRAHSEVOT 589
Qy 553 RIAKIIREKLGKTNADGLTNLTDLNKAFTLYNODMNATDKVYYPNNLPMAYAWMLONK 612
Db 590 VIAKIIKAQI--NPKTDGLT--FTLDELKQAFKIYNEDMQAKKYYQTSNIPTAYALMSLNK 647
Qy 613 DTVTRVYGDYMDNGQYMATKTPFYNAIETLLKGRIKYVAGQ--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYDDQGYMATKSPPYDAITDLLKARIKYAAGGQDMKITTYVEGDKSHMDW 707
Qy 665 S-SGLTISVRVKGANSASDAGNTETNRSGMALLINRPNFRAYRNLTLNMGAAHKSQ 721
Db 708 DYTGLTISVRVYGTGANEATDQGEATKQGMNAVITSNRPNLSKLNQNDKIVNMGAHKNQ 767
Qy 722 AYRPLLLSTKQGIATYLNDSVDSDQYKYTDSQGNLSFASASELQSVANAQVSGMIQVWV 781
Db 768 EYRPLLLSTKQGIATYLNDSVDSDQYKYTDSQGNLSFASASELQSVANAQVSGMIQVWV 826
Qy 782 VGAADNDQVRTSPSTQATKDGNIHQSDALDSQVIEGFSNFOAFAQSPDOVTNAVIK 841

```
Db 827 VGASNDQVRVAASNKANATQVYBSSSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQN 886
Qy 842 GDLFKSWGITQFEMAPQYVSSSDGTFGLDSVLINGYAFSDRYDLAMSKNNKYSGKODLANA 901
Db 887 VQLFKSWGVTSEMAPQYVSSSDGTFGLDSIIQNGYAFEDRYDLAMSKNNKYSGQDDMINA 946
Qy 902 IKGLOSAGIKVLSLVPNQLYNLPKGEVVTATRVNQYQGAQSGATINKTPYVANTRSYG- 960
Db 947 VVALHKSQIQTADWPDQIYNLPKGEVVTATRVNDYGEYKDSSEIKNTLYAANTKSNKG 1006
Qy 961 DYQEOYGGKFLDLDLQKLYPRLFSTQISTGKIDPDSVKITNWSAKYFNGSNILGRGAKV 1020
Db 1007 DYQAKYGGAFSELAAKYPSIFNRTOISNGKKIDSEKITAMKAKYFNGTNILGRGVY 1066
Qy 1021 LSE--GNKYLNLADGKFLPTVLNNTYQPOVQVANGFISKNGGIHYLDKNGQEVKNRP-K 1077
Db 1067 LKDNASDKYFELKGNQYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
Qy 1078 EISGSWYFSDSGKMATGKTKIGNDTYLPMPNGKQLKEGVW--YDGKAYYYDDNGRTWT 1135
Db 1122 DAKGNWYFDNNGHMYVGLQQLNGEVQYFSLNSGVQLRESFLENADGSK-NYFCHLGNRYS 1180
Qy 1136 NKGFEFRVDGDKWRYFNGDGTIAIGVSLDNRTLYFDAYGYQVKGQTVT-INGKSYTF 1194
Db 1181 N-GYYSF--DNDSKWRYFDASGVMAVGLXTINGTQYFDQDGYQVKGAWITGSDGKKYF 1237
Qy 1195 DADQGLVQTDNANPAPOQAGWKLLGDNQWGY-RKDGQLLTGEOITDQKVFQDNGVQ 1253
Db 1238 DDGSGNMAVRPANDK-----NGDWYLLNSDGIALVGVTINGKTYFYFGDQKQ 1286
Qy 1254 VKGGTATDASGLRFDYDRDQHQVKGWYSTSDNNWYVYNGSGOVLTLGTTIDGQTVYFD 1313
Db 1287 IKGKIITD--NGKLYFLANSGLARNIFATDSQNNWYFPGSDGVAVTGQTIAAGKLYPA 1345
Qy 1314 DKGIOAKGKAVDENGRLRYFDADSGNMLDRWK-NVDGNWYFNRNGLA 1362
Db 1346 SDGKQVKGsfv-TYNGKVHYHADSGELQVNRFEADKOGNWYLLDSNGEA 1394

RESULT 4
US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-36

Query Match 43.9%; Score 3174; DB 4; Length 1430;
Best Local Similarity 47.3%; Pred. No. 1.7e-182;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

Qy 1 MEKNRLYKHKVKKQWALGVTT--VTLSEFLAGGQVVAADTNNDGTSV-----QVNRKV 53
Db 1 METKRYKMHKVKGVHVTVAAGLITLGTTLIGSSVSAETEQTSDDKVTQKSEDDKAA 60
Qy 54 PSDPKFDA----QAQNGQLAQMFAKAAQADOTATSOVSPATDGRVDNQTVAAPAAAN 109
Db 61 SESSQTDAPKTKQAQTEQ-TQAQ-SQANVAD-TSIS-----ITKETPSQNTTQANSDDKT 113
```

```
Qy 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDDPG-----H 149
Db 114 VTNTKSEBAQTSEERTKQSEEAQTASSQALTQAKAELTKQRTAAQENKNPVDLAAIPN 173
Qy 150 LETVDGKYYVDANGORLKNYSWVIDGKTYFYDGTGGAOTDLP---KTGQANODNVPD 205
Db 174 VQIDGKYYIGSDGQPKNFALTVNNKVLVFDKNTG-ALTDTSQYQFKGLTKLN--- 229
Qy 206 SYQANNQAYSNEASFETVDNYLTADSWYRPRKILKNGSQWASSEGDLRPLMTWMPDA 265
Db 230 DYTPHNQIVNFENTSLIEDNYVTADSWYRPRKILKNGKTWTASSEDLRPLMLSWMPDK 289
Qy 266 ATKAAYANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEKKITASEGNTWMLDKMSQ 322
Db 290 QTQIAYLYNMQQGLGTGENTYTDSSQESLMLAAQTVOVKIBTKISQTOQTQWLDRDINS 349
Qy 323 FVKSONOWS IASENVTVPNODHMOGGALLFESNKSOTEHANSDMRLLNRPFTQTGKQKY 382
Db 350 FVKTPNWNQSQTESDTSAGEKDLQGGALLYNSDKTAYANSDYLLNRTPTFSQTGKPKY 409
Qy 383 FTTNYA-GYELLANDVDNSNPVQAEQLNHLHYLLNMNGDIYMGDKDANFDGVRVDVADN 441
Db 410 FEDNSSGGYDFLANDIDNSNPVQAEQLNHLHYLLMNYGSI VANDPEANFDGVRVDVADN 469
Qy 442 VNADLLQIORDYYKAKYGTQNEKVAIDHLSILEAWSGNDNDYVQKNNFSLSIDNDQRS 501
Db 470 VNADLLQIASDYLKAHYGVDSKNAINHLSILEAWSNDNDPOYNKDTTGAQLPIDNKLRL 529
Qy 502 GMLKAF-----GYASAYRGNLSNLATAGLKNRSA--NPDSDPVPNTVYFIRAHSEVOT 552
Db 530 SLLYALTPLPEKADSNKNEIRSGLEPVITNSLNRSAGKNSERMANIYFIRAHSEVOT 589
Qy 553 RIAKIIREKLGKTNADGLTNLTDDLKNAFDIYNODMNAATDKVYYPNNLPMAWMLQNK 612
Db 590 VIAKIICAQI-NPKTDGLT-FTLDELKQAFKIYNEQMRQAKKKTQSNIPTAYALMLSNK 647
Qy 613 DTVTRVYGDVMTDNGQWMAKTPPYNAIETLLKRIKRYVAGGQ--AVSYKQ-----DW 664
Db 648 DSITRLYYGDMYSDDGQYMATKSPYDAIDTLTKARIKYAAGGQDMKITYVEGDSKSHMDW 707
Qy 665 S-SGILTSVRYCKGANSADAGNTRNSGALLNNRPNPRAYRN--LTLNMGAAHKSQ 721
Db 708 DYTGVLTSVRYGTGANEATDQSEATKQGMVITSNPNPSLKNQNDKVI VNGGAHAKRQ 767
Qy 722 AYRPLLSTKDGIAIYLNDSVDVSRQYKYTDSQGNLSFSASELQSVANAQVSGMIQVWVP 781
Db 768 EYRPLLTTKDLGTSYTSDAAKS-LYRKTNDKGELVFDASDIQGYLNPQVSGYLAVWVP 826
Qy 782 VQADNQDVRTSPSTQATKQGNIIHQSDALDSQVIYEGFSNFOAFAQSPDQTNVIAKX 841
Db 827 VGASDNQVRVAASNKANATQVYBSSSALDSQLIYEGFSNFQDFVTKDSYTNKKIAQN 886
Qy 842 GDLFKSWGITQFEMAPQYVSSSDGTFGLDSVLINGYAFSDRYDLAMSKNNKYSGKODLANA 901
Db 887 VOLFKSWGVTSEMAPQYVSSSDGTFGLDSIIQNGYAFEDRYDLAMSKNNKYSGQDDMINA 946
Qy 902 IKGLOSAGIKVLSLVPNQLYNLPKGEVVTATRVNQYQGAQSGATINKTPYVANTRSYG- 960
Db 947 VVALHKSQIQTADWPDQIYNLPKGEVVTATRVNDYGEYKDSSEIKNTLYAANTKSNKG 1006
Qy 961 DYQEOYGGKFLDLDLQKLYPRLFSTQISTGKIDPDSVKITNWSAKYFNGSNILGRGAKV 1020
Db 1007 DYQAKYGGAFSELAAKYPSIFNRTOISNGKKIDSEKITAMKAKYFNGTNILGRGVY 1066
Qy 1021 LSE--GNKYLNLADGKFLPTVLNNTYQPOVQVANGFISKNGGIHYLDKNGQEVKNRP-K 1077
Db 1067 LKDNASDKYFELKGNQYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
Qy 1078 EISGSWYFSDSGKMATGKTKIGNDTYLPMPNGKQLKEGVW--YDGKAYYYDDNGRTWT 1135
Db 1122 DAKGNWYFDNNGHMYVGLQQLNGEVQYFSLNSGVQLRESFLENADGSK-NYFCHLGNRYS 1180
```

QY 1136 NKGFVBRVQDKWRYFNGDGTIAIGLSLDRNTLYFDAYGVQVKGQVTV-INGKSYTF 1194
 Db 1181 N-GYISF--DNDSKWRVYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITSGDKKRYF 1237
 QY 1195 DADGDLVQTDNANPAPQAGWKLJLGDNONGY-RKDGLLTGBOITDQKVFQDNGVQ 1253
 Db 1238 DGSNGMAYNRPANDK-----NGDWYILNSDGIALVGQTINGKTYFFGQDGRQ 1286
 QY 1254 VKGGTATDASGLRFLFYDRDQGHQVGKGYSTSDNNWYVNESGQVLTGLQTDGQTVYFD 1313
 Db 1287 IKGKIITD-NGKLKYLPLANSBELARNIFATDSQNNWYVFGSDGVAVTGSQTIAGKLYFA 1345
 QY 1314 DKGIQAGKAVDENGRLRYFDADSGNMLDRWK-NVDGNWYVFNRLGLA 1362
 Db 1346 SDGQVKGFSV-TYNGKVHYTHADSGELQVNRFEADKDGWYVLDNSGEA 1394
 RESULT 5
 US-10-797-821-36
 ; Sequence 36, Application US/10797821
 ; Publication No. US20050031633A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Daniel J.
 ; APPLICANT: Taubman, Martin A.
 ; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
 ; FILE REFERENCE: 25669-020
 ; CURRENT APPLICATION NUMBER: US/10/797,821
 ; PRIOR FILING DATE: 2004-03-09
 ; PRIOR FILING DATE: 2003-03-07
 ; PRIOR FILING DATE: 2002-03-07
 ; PRIOR FILING DATE: 2002-03-07
 ; PRIOR FILING DATE: 2002-08-08
 ; PRIOR FILING DATE: 1999-04-12
 ; PRIOR FILING DATE: 1998-04-13
 ; PRIOR FILING DATE: 1998-04-13
 ; PRIOR FILING DATE: 1999-01-08
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 36
 ; LENGTH: 1430
 ; TYPE: PRT
 ; ORGANISM: Streptococcus mutans
 US-10-797-821-36
 Query Match 43.9%; Score 3174; DB 5; Length 1430;
 Best Local Similarity 47.3%; Pred. No. 1.7e-182;
 Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;
 QY 1 MEKNRLYKLHKYKQWVAIGVTV--VTLSFAGGQVVAADTNNDGTSV-----QVKNV 53
 Db 1 METKRYKMKHKKHKKHVVTVAVASGLITLGTTLGSSVSAETEQQTSKDVVTKSEDDKAA 60
 QY 54 PSDPKFDA---QAQNGQLAQAMFKAAQADOTATSOVSPATDGRVDNVTAAQNPAA 109
 Db 61 SESSQTDAPKTKQAQTEQ-TQAQ-SQANVAD-TSTST-----ITKETPSQNIITQANSDDKT 113
 QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDPG-----H 149
 Db 114 VTNKSEEAQTSERTKQSEEAQTASSQALTAQKAEKLTQRTAQENKNPVDLAAIPN 173
 QY 150 LETVDGKTYVDANGRLKNSWIDGKTYFFDQGTGEAQTDL P-----KYGANQDNVDP 205
 Db 174 VKQIDGKYYIYIGSDGQPKNFALTVNNKVLVDFDKNTG-ALTDTSQYQFKQGLTKLNN--- 229
 QY 206 SYQANNOAYSNRASSPETVDNLTADSWYRPRKILKNGOSMOASSRGLRPLILMTWPD 265
 Db 230 DTPRHQIVNFENTSLTIDNTVADSWYRPRKILKNGKTWASSRDLRPLILMTWPD 289
 QY 266 ATKAAAYANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEKKTIASEGNTNWLDRKMSQ 322

Db 290 QTQIAYLNMNQQGLGTGENYTTADSSQESLNLAAQTVOVKIETKISQTQOTQWLRLDIINS 349
 QY 323 FVKSQNMWSTASNETVYPNQDHWOGALLFNSKQDTEHANSWRLLNRNRPRTQTGQKY 382
 Db 350 FVKTQPNWNSQTESDTSAGEKDHQGLGALLYSNDSKTAYANSYRLLNRPRTQTGPKY 409
 QY 383 FTTNYA-GYELLILANDVNSPNVVAEQALNHLHYLMNWGDI VMGDKDANFDGVRVDAVDN 441
 Db 410 FEDNSSGGVDFLLANDIDNSPNVVAEQALNHLHYLMNYSIVANDFEANFDGVRVDAVDN 469
 QY 442 VNADLLQIORDYKAKYGTQDQNEKNAIDHLSILEAWSGNDNDYVKQNNFSLSDNDORS 501
 Db 470 VNADLLQIASDYLKAHYGVDKSEKNAIHLISILEAWSNDNDPQYNKDTKGAQLPLDNKRL 529
 QY 502 GMLKAF-----GYASAYRGNLSNLATAGLNRSN-NPDSDPVPNRYFIRAHDSVOT 552
 Db 530 SLLYALTRPLEKDASNKNEIRSGLEPVTNSLNRSAEGKNSRMANIYIFIRAHDSVOT 589
 QY 553 RIAKIIREKLGKTNADGLTNLTLDDLKAFDIYNQDMNATDKVYYPNNLPMAYAMLQNK 612
 Db 590 VIAKIIKAOI-NPKTDGLT-FTLDELKQAFKIYNEDMRQAKKYTOSNIPTAFALMLSNK 647
 QY 613 DTVRVYGYDMYTDNQYMATKTPFYNAIETLLKGRIKYVAGQ--AVSYKQ-----DW 664
 Db 648 DSTIRLYGDMYSDDCQYMATKSPYYDAIDTLIKARIKYAAGQDMKITYVEGDKSHMDW 707
 QY 665 S-SGILTSVRYGKANSASDAGNTETNSGMALLNNRNFRAYRN--LTLANGRAHKSQ 721
 Db 708 DYTGLVTSVRYGTGANEATDQSEATKTQGMVITSNNPSLKLNDQDKVIVNNGAAHKNQ 767
 QY 722 AYRPLLLSTKDGITATYLNDSVDROYKYVTDSONLSFSAELQSVANAAQVSMIOVWVP 781
 Db 768 EYRPLLLTTKDGITATYLNDSVDROYKYVTDSONLSFSAELQSVANAAQVSMIOVWVP 826
 QY 782 VQAADNQDVRTSPSTQATKDGNIYHOSDALDSQVIYEGFSNFOAFAQSPDQYTNVIAKN 841
 Db 827 VGASDNQDVRVAASNKANATQVYESSALDSQLIYEGFSNFOAFAQSPDQYTNVIAKN 886
 QY 842 GDLFKSWGITQEMAPQYUSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYSGKQDLANA 901
 Db 887 VOLFKSWGVTSPFEMAPQYVSSDGTFLDSVILNGYAFSDRYDLAMSKNNKYSGKQDLANA 946
 QY 902 IKGLQSAKIKVLSDLVPLNQLYNLPGKEVVTATRVNOYGOAKSGATINKTPYVANTRSYG- 960
 Db 947 VKALHKSGLQVADWVFDQIYNLPGKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKG 1006
 QY 961 DYQBYGKFLDLDLQKLYPRLFTSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
 Db 1007 DYQBYGKFLDLDLQKLYPRLFTSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1066
 QY 1021 LSE--GNKYLNLADGKFLPVTVLNNTYGOVPSANGFISKNGGIHYLDKNGQVQRFP-K 1077
 Db 1067 LKDNASDKYFELKGNQYTLPKQMTN-----KEASTGVNDGNGMTFYSTSGYQAKNSFVQ 1121
 QY 1078 ETSQSWYVPSDQKMATGKTKIGNDTYLPMPNGKOLKEGV--YDCKKAYVYDDNGRTWT 1135
 Db 1122 DAGKNWYVFDNNGHVMYVGLQNGEVOYFLSNGVQLRESFLENADGSK-NYFHLGNRYS 1180
 QY 1136 NKGFVEFRVDGQKWRYPNGDGTIAIGLSLDRNTLYFDAYGVQVKGQVTV-INGKSYTF 1194
 Db 1181 N-GYISF--DNDSKWRVYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITSGDKKRYF 1237
 QY 1195 DADGDLVQTDNANPAPQAGWKLJLGDNONGY-RKDGLLTGBOITDQKVFQDNGVQ 1253
 Db 1238 DGSNGMAYNRPANDK-----NGDWYILNSDGIALVGQTINGKTYFFGQDGRQ 1286
 QY 1254 VKGGTATDASGLRFLFYDRDQGHQVGKGYSTSDNNWYVNESGQVLTGLQTDGQTVYFD 1313
 Db 1287 IKGKIITD-NGKLKYLPLANSBELARNIFATDSQNNWYVFGSDGVAVTGSQTIAGKLYFA 1345
 QY 1314 DKGIQAGKAVDENGRLRYFDADSGNMLDRWK-NVDGNWYVFNRLGLA 1362


```
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
; US-10-797-821-39

Query Match      42.6%; Score 3080.5; DB 5; Length 1554;
Best Local Similarity 43.9%; Pred. No. 8.4e-177;
Matches 667; Conservative 213; Mismatches 446; Indels 193; Gaps 34;

QY 1 MEKRLRYKLVKVKQWVIGVTTVTLTSLFAGGQVAA-----DTNNNG 44
Db 1 MEKRLRYKLVKVKWVTVIAVSLVGLVAGCTVSAEDKAVANDTTAQATVGVDTGDDQA 60
QY 45 TSVQVKNWVPSDPKFDQAQNGLOAQMFKAAQADQ-----TATSQVSPATDGRVDNQVT 100
Db 61 TTNDANTNTTDTADQSAANTNG-DQAGSDQSNQDQAKQDPTANTDRQADNSQTDN--- 116
QY 101 PAANQPAANVANQDANPANTDAGALNRQSAADTST----DGKAVPQTSDDPQ-----HLE 151
Db 117 -----NQATDQATSPATDGTSVQRDAANVATAADQEGQTAPSEQEKSAALSLDNVK 168
QY 152 TVDGKTYVYDANGQRLKNSMVIDGKTYFYDGGTGEAQTDLPKTQGANQDNVPSYQANN 211
Db 169 LIDGKTYVYQADGYSYKKNFAITVNGQMLFYDSGTGALSSTSTVSFSQGTTLNVDDPSSH 228
QY 212 QAYSNEASSFETVDNLTADSVYRPRKILKNGSQWASSEGDLRPLMTLWPDATKAAY 271
Db 229 KAYDSTAKGFELVNGYLTANSWYRPGAILRNGQTWEASNDLRPLVMSWMPDKQTQVAY 288
QY 272 ANFW-----AKEGLISSYRQNSANLDAATQNIQSAIEKKIASSEGNLNRDKMSOFVKS 326
Db 289 VNYMNYKLSANETEVNTESQVDLNKEA--QSIQTKEQKIISDNTQWLRTAMEAFVAA 346
QY 327 QNQWSTASNETVYPNQDHQGALLFSNSKOTEHANSWRLNRNPTFTQTKQKYFTT- 385
Db 347 QPKWNNMSTEN---FNKGDHLQGGALLYTNSDLTPWANSDYRLNRTPTQDGTKKYFTTEG 403
QY 386 NYAGYELLILANDVDSNPVQAEQLNHLHYLNWGDIVNGDKDANFDGVRVDAVDNVDNAD 445
Db 404 GEGGYEFLLSNDVDSNPVQAEQLNHLHYLNWGDIVNGDKDANFDGVRVDAVDNVDNAD 463
QY 446 LIQIQORDYKAKYGTQDNEKNAIDHLSILEAWSGNDNDVYKDONNPSLSIDNDQSGMLK 505
Db 464 LLQVSNYFNKYNKYVTDSEANALAHISILEASLNDQNYEDNTNGTALSIDNSSRLTSLA 523
QY 506 AFGYASAYRGNISNLATAGLKNRSANPDS--DPVPNYVFIKRAHDSVQTRIARIKELG 563
Db 524 VLTKQPGQRIDLSNLISSESVKNERANDTAYGTIPTYSFVRAHDSVQTVIAKIVKEKI- 582
QY 564 KTNADGLTNLTLDLNLKAFDIYNQDNVATDKVYPPNPLPMAYAMLMQNDKTVTRVYGD 623
Db 583 DTNSDGYT--FTLDQLKDAFKIYNEQMAKNTKTYTHNIPAAVALLSNMESVPRVYVYGL 641
QY 624 YTDNGQYMAKTPFYNAIEFLKGRITKYVAGQAVS-YKQDMSGGLTTSVRYGKGANSAS 682
Db 642 YDDQYMAKGPYDAIATMLQGRJAYVSGGQSEVHVKNQNLSSVRYGQDUMSAD 701
QY 683 DAGNTE-TRNSGWALLINNRPNF-RAYRNLTLNMGAAHKSQAYRPLLLSTKDGATYIND 740
```

```
Db 702 DTQGTDLSTRTSGLVTLVSNPDLNLDLGGDSLTVNMGRAHANQAYRPLILGTGKGVSQSLKD 761
QY 741 SDVDSRQYKYTDSQGNLSFSAQSLOVANAQVSGMQLVAVPVCAANDQVDTSPSTQATK 800
Db 762 SDTNI--VKYTDANGNLTTADDIKGYSTVDMSGYLAVVPVGAQKQDQVVRVADTNQKA 819
QY 801 DGNHYHQSALDSQVIYEGFSNFQAFAPQSPDQVTVNAVIAKNGDLFKFSWGITQFEMAPQVY 860
Db 820 DGKSLKTSALDSQVIYEGFSNFQDPANNDADYTNKKIAENADFFKGLGITSFEMAPQVY 879
QY 861 SSEDGTFILSVILNGVAFSDRYDLAMSKNKGYSKODLANA IKGLQSAGIKVLSVLDVNPQ 920
Db 880 SATDGSFLDSIIQNGYAFSDRYDLAMSKNKGYSKODLANAL KALHANGIQALADWVPDQ 939
QY 921 LYNLPGEVVTATRVNOYGOAKSGATINKTPPVANTRSYG-DYQEOYGGKFLDLDLQKLYP 979
Db 940 IYQLPGEVVTATRVNOYGOAKSGATINKTPPVANTRSYG-DYQEOYGGKFLDLDLQKLYP 999
QY 980 RLFTSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSSEG--NKYLNILADGKFL 1037
Db 1000 DMFTVMNISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSSEG--NKYLNILADGKFL 1059
QY 1038 PTVLNNITYCQPOVSANGFISKNGGIHYLDKNGQEVKNRKEISGSIYFSDSGKMATG-- 1095
Db 1060 PA-----SFTGCDQNAKTGFYDGTGMAYYSTSGNKAVNSFIYEGCHYVYFDKDHMTG 1115
QY 1096 KTKIGNDTYLPWPNKGQKLEGVWYDCK-KAYYYDDNNGRTWTKNG-----FVEFRVDGQDK 1149
Db 1116 KAEDGND-VYFLPNGIQMRDAIYQDAQNSYVYGRGTGILY--KGDNWPVFPVDPNNANKTV 1172
QY 1150 WRYFNGDGTIAICLSLDNRITLYFAYGYQVKGQTVITNGKSYTFDADQG----- 1199
Db 1173 FRYFDANNVWALGYRNMYGQTYVYFDENGFOAKGQLLTDKGTHTYFDEDNGAMAKNFVNV 1232
QY 1200 --DLVQTDNANPAQOQ----- 1214
Db 1233 GDDWYVMDGNGNAVKGOYVNNQILYFNPETGVQVKGQFITDAQGRTSYVDANSALKSS 1292
QY 1215 -----AGMKLLGDN--OHGY--RKDQQLLTGQDTIDGQKVFPQ-DN 1250
Db 1293 GFTFNGSDWYVAENGYVYKGFKAENQDQYFDQTTGKQAKGAQKADGRDLYFNPDS 1352
QY 1251 GYQVKGKGTATDASGVFLRYDQHQGVKGWSTSDNNWYVYVNESQVLTGLQTIID---- 1306
Db 1353 GYQVKGKGFATDESNTSFVHGDNKGDKVCGFFTTGNNAYYADNNGNLVKGFEIDGKWY 1412
QY 1307 -----GQTVYFD-DKGIQAKGKAVMDENGNLRYPDADSGNMLRDRWKN 1348
Db 1413 HFDEVTTQQAAGALVNGQQLYFDVDSGIVQVKGDFVTDGQNTSYVDVNSGD-----KK 1466
QY 1349 VDG-----NWYFNRNG 1360
Db 1467 VNGFPTTGDNAWYADGQG 1485
```

RESULT 8

```
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
```

```
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match          41.9%; Score 3031; DB 4; Length 1590;
Best Local Similarity 41.8%; Pred. No. 8.4e-174;
Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;

QY 1 MEKNRYKLHKVKQKOWVAIGVTVT--LSFLAGQVVAADTNNNDGTSVQNNKVPSPDKPF 59
DB 1 MEKNVFPKHVKERWVTLVSASATMLASALGASVASAUTDTASDDSNQA--VVTGD--- 55
QY 60 DAQAQNGQLAAQAFKAANOADQATTSQVSPATDGRVDNQVTPAANPAANVANQDV---- 115
DB 56 --QTTNNQATDQTSIAA-----TATSEQASASTDAATDQ--ASAAEQTQGTASTDTAAQT 106
QY 116 ---ANPA-----TDAGALNRQSAADTSTDGKAVPQTSDDPGHLETVDGKTYV 160
DB 107 TTNANEAKWPTENENOGFTDEMLAEAKNVATAESD--SIPSDLAKKSNVKQVDGKIYY 164
QY 161 DANGQRLKKNYSMDVCKTYFQGTG---EAQTDLPKTKQANQDNVPDSYQANNOAYSN 216
DB 165 DQDGNVKKNFVSVGDKIYYFD-ETGAYKDTSKVDADKSSAVSQA-TIFAANPEAYST 222
QY 217 EASSPETVDNYLTADSWYRPRKILKNGSQWASSEBGLRPLILMTWPPDAATKAAYANFWA 276
DB 223 SAKNFEAVDNYLTADSWYRPKSILKDKTWTESGDDPRPLLMWMPDTETKRYNYVYN 282
QY 277 KEGLISGSY--RQNSANLDAATONIOSAEIKKTIASEGNTNWLDRKMSQFVKSONOWSIAS 334
DB 283 KVVGDIKTYTAETSQADLTAAAEVLQARIEQKITSNNNTKWLREATSAFVKTPQPMNGES 342
QY 335 ENETVPNODHMOGGALLFSNSKD-TEHANSWRLLNRNPTFOTGK-QKYFTTN-----YA 388
DB 343 EK-----PYDDHLQNGALLFDNQDLDTPDQSNVRLNRTPTNQTSGLDSRFTYNPNDPLG 398
QY 389 GYELLANDVNSNPVQAEQLNHLHYLMNMGDIVNGDKDANFPDGVYRVDAVNVDNADLLQ 448
DB 399 GYDFLLANDVNSNPVQAEQLNHLHYLMNFGSIYANDADANPDSIRVDADVNDVADLLQ 458
QY 449 IQRDYVYKAKYGTONEKNAIDHLSILEAMSGNDNDVVKDQNNFSLSDNDQSGMLKAFG 508
DB 459 ISSDYLKAAAYGIDKNNKNANNHVSIVEAWSDDNTPYLHDDGDNLMNMDNKRFLSLMSLSLA 518
QY 509 YASAYRGNLSNLATAGLKNRSANP-DSDDVPNVPYFIRAHDSVEQTRIAKIIREKLGKTN 567
DB 519 KPLDKESGLNPLIHNLSLVREVDREVEITVPSYSFARAHDSVEQDIIRDIKAEI-NPRS 577
QY 568 DGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAYAMWLQNKDVTTRVYVYGMWYDN 627
DB 578 FGYSFTQBEIEQAFKIYNEDLKKTKYTHVNPVLSYLLTLTKGSIPIRVYVYGMFTDD 636
QY 628 GQYMAKTFPYNAIETLLKRIKIVAGQAVSYKQDMSSGILTSVRYGKGANSASDAGNT 687
DB 637 GQYMAKNTVNYDAIESLLKARKMYGSGQAMQNYQIENGEEILTSVRYGKGALKQSDKGBA 696
QY 688 ETRNSGALLINRNPFR- YRNLNLMGAAHKSQAYRPLLSTKDGIAFYLNDSVD--S 745
DB 697 TTRTSGVGVMGQNPFLDGKVALNMGNANQYERALMVSTKDGVIATYATDASKA 756
QY 746 ROYKYTDSQNLFSASEIQSVAANAQVSGMIQWVPVGAADNQDVRTSPSTOATKDNITY 805
DB 757 GLVKRTDENGILYFLNDDLKGVANPQVSGFLQWVPVGAADDQDIRVAASDASTADGKSL 816
QY 806 HQSDALDSQVIVGFSNFQAQSPPOYTNNAVITAKNGDLFKSGIITQFEMAPQYVSSG 865
DB 817 HODAAADS RVMPFGFSNFQSFATKEBEYNNVJANNVDFSVWGIITDFEMAPQYVSSDTG 876
QY 866 TFLDSVILNGYAFSDRYDLAMSNNKYGSKQDILANAIGLQSAIGIKVLSDLVNPQNLNLP 925
DB 877 QFLDSVJONGYATTDYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPQMYTFP 936

QY 926 GKEVVTATRVNQYQAKSGATINKTPYVANTPSYG-DYOEQYGGKFLDLDLQKLYRPLPST 984
DB 937 KOEVVTVTRTKDFGKPIAGSQINHLSTYVDTKSSGDDYQAKYGGAPLDELKPKYELPFTK 996
QY 985 KQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKTVLSE--GNKYLNILADGKLFPTVLN 1042
DB 997 KQISTGOALDPSVKIKQWSAKYFNGSNILGRGADVLSDOVSNKYFNVASDTLFLPSSLL 1056
QY 1043 NTYGQPVSAVANGFISKNGIHYLDKNGQEVKMRPKIEGSMWYFSDSGMATGKTIGND 1102
DB 1057 GKVESGIRYDG---KGYIYNSSATGQVKASFITEAGNLYYFGDGYMTVGAQTINGA 1112
QY 1103 TYLFNPNKQLKEGVWYDGK-KAYYYDDNGRWTNKGFPVPRVDGQKWRYNFGDGTIAI 1161
DB 1113 NYFFLENGTALRNTIYTDAGNSHYIYANDGRYEN-GYQOF---GND-WRYPK-DGNMAV 1166
QY 1162 GLVSLD----- 1167
DB 1167 GLTTVDGNVOYFDPKDGVOAKDKLIIVTRDGKVRYPFHNGNAATNTFIADKTGHWYLGKD 1226
QY 1168 -----NRTLYFDAYGYQVKGQTVTIN-GKSYTFDADQGD----- 1201
DB 1227 GVAVTGAQTVGKQKLYFEANGQOVKGDFTVSDGKLYFYDVSDDMTDTFTIEDKAGNWF 1286
QY 1202 ----- 1201
DB 1287 YLKGDAAVTGAQTIRGQKLYFKANGQOVKGDIVKGTGKIRYDAKSGEQVFNKTVKAA 1346
QY 1202 -----VOTDNNANPAP-----OQO-----AGWKLIGDNQWGYRKDG 1231
DB 1347 DGKTVYVINDGVAVDPSPVVKGTQPKDASGALRFPYLNKGLVGTSGWYETANHDWYVIOG 1406
QY 1232 QLLTGEQITDQKVPFQDNGVQVKGATATDASGLVLFYDRDQGHQ----- 1276
DB 1407 KALTGEQITNGHLYFKEDGHQVQLVGTGTDKVRYYDANSQDQAFNKSVTVNGKTYVF 1466
QY 1277 -----VKGWYSTSDNNWVYVNESGOVLTLGQ 1303
DB 1467 GNDGTAQTAGNPKGTQDKGSDIRFYSMEGQVLTSGWYENAGQWLYV-KNGKVLTLGQ 1525
QY 1304 TIDGQTVYFDDKGIQAKKAVMDENGLRYFDADSGNMLRDRWKNVNDGNWYVFNRRNG 1360
DB 1526 TVGSRVYFDENGIQAKKAVRTSGKIRYFDENSGSMITNQWKFYGYGYFFGNDG 1582

RESULT 9
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; PRIOR FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
```

```
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match      41.9%  Score 3031; DB 5; Length 1590;
Best Local Similarity 41.6%  Pred. No. 8.4e-174;
Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;

QY 1 MEKNLYKHLKVKQWVAIGVTTVT-LSPFAGGQVVAADTNNNDGTSVQVKNKVPDPKF 59
DQ 1 MEKNVAFQHKVKKRWVTLVASATMLASALGASVASADTDTASDSDNQA--VVTGD--- 55

QY 60 DAQAQNGOLAQAAMFKAANQADQATSOVSPATGRVDNQVTPAANQPAANVADV--- 115
DQ 56 -QTTNQADTQTSIAA-----TATSEQASASTDAATDQ--ASAAEQGTGTASTDTAAQT 106

QY 116 ---ANPA-----TDGALNRQSAADTSTDGKAVPQTSDDPGHLETVDGKTYV 160
DQ 107 TTNAEAKVPTENENQGFDEMLAEAKVATAESD--SIPSLAKSNVQVQDGKYYY 164

QY 161 DANGORLKNYSMVIDGKTYFFDQGTG----EAQTDLPKTAQANQDNVPSQYANNAISN 216
DQ 165 DQGNVKNQFAVSGVKIYFDD-ETGAYKDTSKVDADKSSAVSQNA-TIFAANNAYST 222

QY 217 EASSFETVDNLTADSWYRPRKILKNGQSQWQASSEGDLRPILMTWPDAAATKAAANFWA 276
DQ 223 SAKNFEAVDNYLTADSWYRPRKSLDKGTWTEGKDDFRPLLMAWPDTETKRNYVNYW 282

QY 277 KEGLISGSY--RONSANLDAATONIOSALEKKTASGNTNWLDRKMSQFVKSONOWSIAS 334
DQ 283 KVGIDKTYTAETSQADLTAAELVQARIEQKTSNNTKWLREASAFVKTPQWNGES 342

QY 335 ENETVYPNODHMGQGALLFSNSKD--TEHANSWRLNRPFTQTKG-QKYFTTN----YA 388
DQ 343 EK----PYDHLQNGALLFDNQDLDTPDTCQSNRYLLNRTPTNQTSLSDRFTYNDPLG 398

QY 389 GYELLANDVNSNPVQAEQLNHLHYLMNWGDIVMGDKDANPDGVYRDAVNVDNADLLQ 448
DQ 399 GYDFLLANDVNSNPVQAEQLNHLHYLMNWGDIVMGDKDANPDGVYRDAVNVDNADLLQ 458

QY 449 IQRDYKAKYGTQDNEKNAIDHLSILEANSNDNDVYKQDNFSLIDNQDSGLMKAPG 508
DQ 459 ISDYLKAAAGIDKNNKANNHVSIVEAWSNDNTPYLHDDGDNLMMDNKKFLSLMLWSLA 518

QY 509 YASVRGNLSNLATAGLNANP--DSDPVPNVFIRAHDSSEVQTRIAKIIREKLGKTA 567
DQ 519 KPLDKESGLNPLHNSLVREVDREVEVPSISFARAHDSVEQDIIIRIKAEI-NPNS 577

QY 568 DGLTNLTLDLNAKAFDIYNQDNMATDKVYYPNNLPMAYAMLQNDKTVTVYVGYDMYDN 627
DQ 578 FGY-SFTQBEIEQAFKYNEDLKTKDKYTHYVPLSYTLTLTNKGSIPRVYVYDGMFTDD 636

QY 628 GOYMATKTPFYNAIETLLKGRIKYVAGGQAVSKQDWSSGILTSVRYGKANSASDAGNT 687
DQ 637 GOYMAKNTVYDAIESLLKARMKYVSGGQAMQNYQIGNGEILTSVRYGKALKQSDKGA 696

QY 688 ETRNSGMLLNNRPNFRA--YRNLTLNMGAAHKSQARYPLLLSTKDGIAATYLNDSVD--S 745
DQ 697 TTRTSGVGVVMGNQPNFSLDGKVALNMGAAHANQYRALMSTKDGVIATYADADASKA 756

QY 746 ROYKYTDSQGNLSFSAEISQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKDNTY 805
DQ 757 GLVKRTDENGVLVFLNDDLKGVANPQVSGFLQVWVPVGAADQDQIRVAASDSTASTDGSL 816

QY 806 HQSDALDSQVIVYEGSNFQAFQSPDYNTNAVIAKNGDLFKSWGITQFEMAFQYVSSSDG 865
DQ 817 HQDAAMDSDRMWFEFSNFQSPFQKEEYTNVVIANNVDVFXVSGIITDFEMAFQYVSSSDG 876

QY 866 TFLDSVILNGYAFSDRYDLWMSKNNKYQSKQDLANAIKGLQAGIKVLSDLVNPOLYNLP 925
DQ 877 QFLDSVILNGYAFSDRYDLWMSKNNKYQSKQDLANAIKGLQAGIKVLSDLVNPOLYNLP 936

QY 926 GKEVVTATRVNQYQAKSGATINKTPYVANTRSYG--DYQEQYGGKFLDLDLQKLYPRLFST 984

; RESULT 10
; US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
```

US-09-740-274-2

ORGANISM: Streptococcus mutans

Query Match 40.7%; Score 2945.5; DB 3; Length 1475;
 Best Local Similarity 45.8%; Pred. No. 1.1e-168;
 Matches 652; Conservative 225; Mismatches 425; Indels 121; Gaps 40;

QY 1 MEKRLRYKLHKYKQWALGVTT--VTLSFLAGGQVVAADTN-----NNDGTSVQVKN 51
 DB 1 MDKKVRYKLKRVKRWVTVSVASAVMTLTTLTSGG-LVKADSNESKQISNDSNTSV---- 55

QY 52 MVSDFPKDPAQAOQOLAQAMEK--AANQADOT-----ATSOVSPTADGRVDNQVTPA 102
 DB 56 -----VTANEESNVITATSKQEAASQTNHTVTTSSSTSVVNPKE--VSNPYT-- 104

QY 103 ANQPAAN---VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDFQGHLET 152
 DB 105 VGETASNGEKLQNTTTVDKTSSEAAANNISKOTTEADTDVDDSSNAANLQILEKLPNVKE 164

QY 153 VDGKTYVVDANGORLKNYSWIDGTYYPDGTGTAQTDLPKTGOANQDNV---PDSYQA 209
 DB 165 IDGKYVYDNNKVRNFTLIADGKILHPD-ETG-AYTD-TSIDTVNKDI1VTRSNLYKK 221

QY 210 NQQAYSNEASSFTVDNYLTADSWYRPXKILKNGOSWQASSEGDLRPILMTWPDAAATKA 269
 DB 222 YNQVYDRSAQSPFHVHYLTAEWSYRPXYLKDGKXTWTQSTKDFRPILMTWPDQETQR 281

QY 270 AYANFAWEKGLISGSY--RONSANLDAATONIQSAIEKKIASSEGNTNWLKDKMSQFVKSQ 327
 DB 282 QYNYNNAQGLINKTYDDTNSQLQNLIAAAT1QAKIEAKITTLKNDWLRQTI1SAFVKQTQ 341

QY 328 NQMSIASENETVYPNODHMQGALLFSN-SKDTTEHANSQWRLNRPNTFOTGKQ--KYFT 384
 DB 342 SAMNSDSEK-----PFDHLQNGAVLYDNEGKLTPTYANSYRILNRTPTNQTKKDPRYTA 397

QY 385 TN-YAGVELLAVDVSNSPVPQAEOLNHLHYLMNWGDI1MGDKDQANDPDGVRVDAVDNVN 443
 DB 398 DNTIGGYEFLLAVDVSNSPVPQAEOLNHLHFLMFGNIYANDPDANFDSIRVDAVDNDV 457

QY 444 ADLLOLORDYKAKYCTDQEKNAIDHLSILEAWSNDNDVVKQONNFSLSIDNDQRSGM 503
 DB 458 ADLLQAGDYLYKAAKGHHKNDKAANDHLSILEAWSNDNDTPYLHDDGNMIMNDKLRUSL 517

QY 504 LKAPGVASAVRGNSLNTAGLKQRS-ANPDSDDPVNPTVFI1RAHDSVETQRIAKI1REKL 562
 DB 518 LPSLAPLNQSRGWNPLI1TNSLVNRTDDNAETAAPVPSYFFIRAHDSVETQDLI1KAEI 577

QY 563 GKTNADGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAAYMLQNKQDVTTRVYVGD 622
 DB 578 -NPNVVGY-SFTMEETKCAFEIYNKOLLATEKXKTHYNTALSYALLTNKSVPRVYVGD 635

QY 623 MYTDNGOYMATKTPFYNA1ETLLKGRIKVVAGQAVSYKQDWSSGILTSVRYGKANSAS 682
 DB 636 MFTDDQYWAHKTI1NTYEA1ETLLKARIKIVSGQAMRNQOVGNSEI1TSVRYGKALKAT 695

QY 683 DAGNTERNSGALLNNRPNER--AYRNLTLNMGAHKSQAYRPLLLSTKDGATYLLND 740
 DB 696 DTGRTTRTSQVAVIEGNPSPSLRKASDRVVVNMGAHKQAYRPELLFTTDNGIKAYHSD 755

QY 741 SDVDSRQYKYTDSQGNLSPSASELOSVANAQVSGMI1QVWVPVGA--DNQD 789
 DB 756 QEA-AGLVR YTNDRGELI1FTAADIKGYANPQVSGYLVGVVVPVGAAL1KMFALRLARPHQ 814

QY 790 VRTSPSTQATKQNI1YHQSALDSQVIYEGPSNFQAFQSPDYTNVIAKNGDLFKSWG 849
 DB 815 MAS-----VHQNAALDSRVYMFEGSNFQAFATKKEEYTNVVIKNDVDFAEWG 862

QY 850 ITQFEMAPQVSSDGTGLDSVLNGYAFSDRYDLAWSKNKYKQKDLANAKGLQSAQ 909
 DB 863 VTDFEMAPQVSSDGSFLDSVLQNGYAF1TRDYLGISKPNKYGTADDLV1KAKLHSEK 922

QY 910 IKVLSLDLVPNLQYLPCKEVVTVATRYNQYGOAKSGATINKTPYVANTRSYG--DYOEQYGG 968

QY 923 IKVMADWVPDQMYAFPEKEVVTVATRVKYGTPVAGSQIKNTLYVVDGKSSGKQQAQYVG 980

QY 969 KFLDDLQKLYPLPSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYTVLSE--GNK 1026
 DB 983 AFLEELQAKYPLPFAKQISTGVPMDPSVKIKOWSAKYFNGTNI1LGRGAGYVLKQDQATNT 1042

QY 1027 YLNLADGK--LPLPTVLTANTYQCPQVSANGFISKNGGHLYLDKNGQEVKRNKFEISGSWY 1084
 DB 1043 YFNISDNKEINPLPKTLN-----QDSQVGFSDYDGKGVYVYSTSGYQAKNTFISEGDKWY 1097

QY 1085 YFDSQKMATGKTKIGNDTYLFMPNGKOLKEGVW--YDGKAYYYVDNNGRTWTNKGVFVEF 1142
 DB 1098 YEDNNGYVTTGAOSINGVNYFSLNGLQRLDAILKNEDGTAYYVND-GRYEN-GYYQF 1155

QY 1143 RVDGQDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTV-TINGKSTVTFDADQDGL 1201
 DB 1156 M---SGWRHFEN-NGEMSVELTVIDGQVYFDEMGYQAKGFVTTADGKIRYFDKQSGNM 1211

QY 1202 VQTDNANPAQCCAG-WKLIGDNQWGYEKDQGLTGTGQTDIGQVFFQDNGVQVKGGTAT 1260
 DB 1212 YR-----NRFIENEKGLYLGE-----DGAAVTGSQTINGQHLHYFRANGVQVKGFBVT 1260

QY 1261 DASGVLYRFYDRDQGHQVKGWYVSTSDNNWVYVNESGQVLTGLQTI1DGQTVVYFDDKGIQAK 1320
 DB 1261 DHGRIISYDNGSGDQIRNFRVNAQGVFFPDNNGYAVTQARTINGQLLYFRANGVQVVK 1320

QY 1321 GKAVDENGRLRYFDADSGNMLRDWR-KXVDGNWYFNRNGLA 1362
 DB 1321 GEFVTDYGRISYDNGSGDQIRNFRVNAQGVFFPDNNGYA 1363

RESULT 11
 US-10-383-930-34
 ; Sequence 34, Application US/10383930
 ; Publication No. US20040127400A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Daniel J
 ; APPLICANT: Taubman, Martin A
 ; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
 ; FILE REFERENCE: 25669-018
 ; CURRENT APPLICATION NUMBER: US/10/383,930
 ; PRIOR FILING DATE: 2003-03-07
 ; PRIOR APPLICATION NUMBER: 60/402,483
 ; PRIOR FILING DATE: 2002-08-08
 ; PRIOR APPLICATION NUMBER: 60/363,209
 ; PRIOR FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 34
 ; LENGTH: 1475
 ; TYPE: PRT
 ; ORGANISM: Streptococcus mutans
 ; US-10-383-930-34

Query Match 40.7%; Score 2945.5; DB 4; Length 1475;
 Best Local Similarity 45.8%; Pred. No. 1.1e-168;
 Matches 652; Conservative 225; Mismatches 425; Indels 121; Gaps 40;

QY 1 MEKRLRYKLHKYKQWALGVTT--VTLSFLAGGQVVAADTN-----NNDGTSVQVKN 51
 DB 1 MDKKVRYKLKRVKRWVTVSVASAVMTLTTLTSGG-LVKADSNESKQISNDSNTSV---- 55

QY 52 MVSDFPKDPAQAOQOLAQAMEK--AANQADOT-----ATSOVSPTADGRVDNQVTPA 102
 DB 56 -----VTANEESNVITATSKQEAASQTNHTVTTSSSTSVVNPKE--VSNPYT-- 104

QY 103 ANQPAAN---VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSDFQGHLET 152
 DB 105 VGETASNGEKLQNTTTVDKTSSEAAANNISKOTTEADTDVDDSSNAANLQILEKLPNVKE 164

QY 153 VDGKTYVVDANGORLKNYSWIDGTYYPDGTGTAQTDLPKTGOANQDNV---PDSYQA 209
 DB 165 IDGKYVYDNNKVRNFTLIADGKILHPD-ETG-AYTD-TSIDTVNKDI1VTRSNLYKK 221

QY 210 NQQAYSNEASSFTVDNYLTADSWYRPXKILKNGOSWQASSEGDLRPILMTWPDAAATKA 269
 DB 222 YNQVYDRSAQSPFHVHYLTAEWSYRPXYLKDGKXTWTQSTKDFRPILMTWPDQETQR 281

QY 270 AYANFAWEKGLISGSY--RONSANLDAATONIQSAIEKKIASSEGNTNWLKDKMSQFVKSQ 327
 DB 282 QYNYNNAQGLINKTYDDTNSQLQNLIAAAT1QAKIEAKITTLKNDWLRQTI1SAFVKQTQ 341

QY 328 NQMSIASENETVYPNODHMQGALLFSN-SKDTTEHANSQWRLNRPNTFOTGKQ--KYFT 384
 DB 342 SAMNSDSEK-----PFDHLQNGAVLYDNEGKLTPTYANSYRILNRTPTNQTKKDPRYTA 397

QY 385 TN-YAGVELLAVDVSNSPVPQAEOLNHLHYLMNWGDI1MGDKDQANDPDGVRVDAVDNVN 443
 DB 398 DNTIGGYEFLLAVDVSNSPVPQAEOLNHLHFLMFGNIYANDPDANFDSIRVDAVDNDV 457

QY 444 ADLLOLORDYKAKYCTDQEKNAIDHLSILEAWSNDNDVVKQONNFSLSIDNDQRSGM 503
 DB 458 ADLLQAGDYLYKAAKGHHKNDKAANDHLSILEAWSNDNDTPYLHDDGNMIMNDKLRUSL 517

QY 504 LKAPGVASAVRGNSLNTAGLKQRS-ANPDSDDPVNPTVFI1RAHDSVETQRIAKI1REKL 562
 DB 518 LPSLAPLNQSRGWNPLI1TNSLVNRTDDNAETAAPVPSYFFIRAHDSVETQDLI1KAEI 577

QY 563 GKTNADGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAAYMLQNKQDVTTRVYVGD 622
 DB 578 -NPNVVGY-SFTMEETKCAFEIYNKOLLATEKXKTHYNTALSYALLTNKSVPRVYVGD 635

QY 623 MYTDNGOYMATKTPFYNA1ETLLKGRIKVVAGQAVSYKQDWSSGILTSVRYGKANSAS 682
 DB 636 MFTDDQYWAHKTI1NTYEA1ETLLKARIKIVSGQ

Db	398	DNTIGGYEFLANDVNSNPVQAEQUNLWHLNFNGNIYANDPANFDSIRDAVDNDVD	457
Qy	444	ADLLQIQORDYYKAYGTDQNEKNAIDHLSILEAWSGNDNDYVKDQNNFSLSDNDQSRGM	503
Db	458	ADLLQIAGDYLKAAGIKHKNDKAANDHLSILEAWSDNDTPYLHDGDNMINMDNKLRLSL	517
Qy	504	LKAFGYSAYRGNLSNLAATAGLKQRS--ANPDSDPVENVYFIRAHDESEVQTRIAKILREKL	562
Db	518	LFLSLAKPLNORSQGNPLITNSLNRDNDNAETAAPVSYSFIRAHDESEVQDLIAIKAEI	577
Qy	563	GKTNADGLTNLTDLLNKAFDIIYNQDMNATDKVYYPNNLPMAIYMWLQNKDITVRVYVGD	622
Db	578	-NPNVGY--SFTWEEIKAEIYNKOLLATEKKYTHYNTALSVALLLTNKSSVPRVYVGD	635
Qy	623	MYTDNGOYMATKTPFYNAIETLLKGRIKYVAGQAVSYKODWSSGILTSVRYKKGANSAS	682
Db	636	MFTDDGOYMAHKTINVEAIEETLLKARIKYVSGQAMRNQOVGNSEIITSVRYKKGALKAT	695
Qy	683	DAGNTEPRNSGMALLINNRNFR--AYRNLTNMGAAHKSOAQRPLLLSTPKDGIATVIND	740
Db	696	DTGDRITRTSGVAIEGNNPSLRKLKASDRVVVNMGAARHKQAQRPLLLITDNGIKAYHSD	755
Qy	741	SDVDSROYKYTDSQGNLSFSAELOSANAAQVSGMIQVWYPVGA--DNQD	789
Db	756	QEA--AGLVRTNDRGELIFTAADIKYANPQVSGYLGWVPVPGAALIKMPALRLARPHQ	814
Qy	790	VRTSPSTQATKGNIIHQSDALDSQVIYEGFSNFOFAQSPDOYTNAVIKNGDLFKSWG	849
Db	815	MAS-----VHQNAAALDSRVNFEFSNFOFAPATKBEYTNVVIKANDYKFAEWG	862
Qy	850	ITPEMAPQVSSGDEGTFDLSVTLNGYAFSDRVDLAMSKNKYGSKODLANAIKGLQSAG	909
Db	863	VTDFEMAPQVSSDTSGFSLSVTLQNGYAFDTRVDLGLSKENKYGTADDLVKAIKALHSK	922
Qy	910	IKVLSDLVNPOLYNLPKQVVTATRVNQYGOAKSGATINTPTVANTRSYG--DYQSOYGG	968
Db	923	IKVMADWVPQMYAFPEKEVVTATRVDKYTPVAGSQIKNTLVVVDGKSSGKQQAQYGG	982
Qy	969	KFLDLDLQKLYPLRSTKQISTGKPIDPSVKITNWSAKYFNNGSNILGRGAKYVLSE--GNK	1026
Db	983	AFLEELQAKYPELPARKQISTGVPMDDPSVKIKQWSAKYFNGTNILGRGAGYVLKQATNT	1042
Qy	1027	YLANLADGK--LFLPTVLNNTYGOQVPSANGFISKNQGIHYLDKNQGEVKNRFEISGSWY	1084
Db	1043	YFNISDNKEINFLPKTLN-----QDSQVGSFYDGKGYVYYSYSGYQAKNTFISEGDKWY	1097
Qy	1085	YFSDGKMACTKTINGDNTFLWPNQKOLKEGVW--YDGKAKYAYDDNGRTWTKGPFVEF	1142
Db	1098	YFONNGYMWYGAOSINGVNTYFISNGIQLDLALKNEDGYIAYYGNDRGRYEN--GYIQF	1155
Qy	1143	RVDQDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTV--TINGKSYTFDADQGD	1201
Db	1156	M---SGVWRHFN--NGEMSVGLTVIDGQVQYFDEMGYQAKGFVTADGKIRYFDKQSGNM	1211
Qy	1202	VQTDNANPACQQAG--WKLLGDQWQYRKDGQLLTGEOITDGGKVPFDQNGVQVKGSTAT	1260
Db	1212	YR-----NRFTIENBEGKLYLGE-----DCAAVTGSQTINGHLYFRANGVQVKGFEVT	1260
Qy	1261	DASGLVRFYDRDQGHQVGKQWYSTDSDNNVYVNESGOVLGTLOTIDQTVYFDDKGIQAK	1320
Db	1261	DHGRISYYDNGSGDQIRNFRVNAQCQWFFYFDNNGYAVTGARTINGQLLYFRANGVOVK	1320
Qy	1321	GKAVWDENGLRYFDDADSGNMLDRW--KNVDGNWYFNRNGLA	1362
Db	1321	GEFVTRDYGRISSYDNGSGDQIRNFRVNAQCQWFFYFDNNGYA	1363

Db 579 PSYFIRAHDSVQDLIRNIIRTEI--NPNVVG--SFTTEIKKAFBIYNKDLATEKYYT 636
Qy 598 PNNLPYAWMLQNKOTVTRVYGYDMYTDNGQYMATKTPFYNAIETLLKGRIKYVAGGOA 657
Db 637 HYNLTALSYALLTNKSSVPRVYGYDMFTDQGYMAHKTINYEALETLLKARIKYVSGGOA 696
Qy 658 VSYKQDSSGILTSVRYGKANSADAGNTETRTNSGMALLINRPNFR--AYRNLTLNMG 715
Db 697 MRNQGVNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNNPSRLKASDRVVVNM 756
Qy 716 AAHKSQAYRPLLSTKDGATYLNDSVDVSRQYKTYDSQGNLSFSAELOSQVANAQVSGM 775
Db 757 AAHKNQAYRPLLSTKDGATYLNDSVDVSRQYKTYDSQGNLSFSAELOSQVANAQVSGM 815
Qy 776 IQVWVPVGAADNQDVRTSPSTQATKDGNIYHQSDALDSQVIYEGFSNFQAFQSPQYTN 835
Db 816 LGVWVPVGAADNQDVRTSPSTQATKDGNIYHQSDALDSQVIYEGFSNFQAFQSPQYTN 875
Qy 836 AVIAKNGDLFKSWGITQFEMAPQYVSSDEGTFLLDSVILNGYAFSDRYDLAMSKNKYGS 895
Db 876 VVIKKNVDFAEWGVTFDEMAPQYVSSDEGTFLLDSVILNGYAFSDRYDLAMSKNKYGS 935
Qy 896 QDLANAIKGLQSAIGIKVLDVNLNPKKEVWTATRVNVOYQKASGATINKTPYVAN 955
Db 936 DDLVKAIKALHSGIKVLDVNLNPKKEVWTATRVNVOYQKASGATINKTPYVAN 995
Qy 956 TRSYG--DYQEYGGKFLDLOKLYPLRSTQKSTGKPIDPSVKITNWSAKYFNGSNILG 1014
Db 996 GKSSGKQQAQYGGAFLEELQAKYPELFARKQISTGVPMDFPSVKIKQWAKYFNGSNILG 1055
Qy 1015 RGAQYVLSB--GNKYNLADGKLFLPTVLNNTYGPQVPSANGFISKNGGIHYLDKNGQEV 1072
Db 1056 RGAGYVLYKQDQATNTYFSLVSDNTFLPKSLVNPNGHSTSSVTGLVFDGKGVYVYSTSGNQA 1115
Qy 1073 KNRKEISGSWYFSDGKWKATKTKIGNDTYLFMPNGKQKEGVYDGGKAY--YDDNG 1131
Db 1116 KNAFTSLGNWYFSDGKWKATKTKIGNDTYLFMPNGKQKEGVYDGGKAY--YDDNG 1175
Qy 1132 RTWTNKGFEVFRVGDQKRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGOTVTINGKS 1191
Db 1176 RYVEN--GYLYF--GQ--QWRYFQ--NGIMAVGLTRVHGAQVYFDASGQKGFQFI----- 1223
Qy 1192 YTFDADQGLVQTDNANPAPQOAGWKLGLDGNQMGYRKDGQLLTGEBQTTIDGQKVFQDNG 1251
Db 1224 ----- 1223
Qy 1252 VQVKGGTATDASGLVRFYDRDOHQVGKGYSTSDNNWYVNESGQVLTGLQTTIDQVY 1311
Db 1224 -----TTADGKLYFDRDSGNQISNRFVRSNGEWFDFHNGVAVTGTVTFFNGRLY 1275
Qy 1312 FDDKGIQAKGAVWDENGLRYFDADSGNMLRDRW--KNVDGNWYFNRNGLA 1362
Db 1276 FKPNGVQAKGEFIRDANGLYRYDPNSGNEVRNFRVRSNGEWFDFHNGIA 1327

RESULT 14

US-10-383-930-35

; Sequence 35, Application US/10383930

; Publication No. US20040127400A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A

; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

; FILE REFERENCE: 25669-018

; CURRENT APPLICATION NUMBER: US/10/383,930

; PRIORITY FILING DATE: 2003-03-07

; PRIORITY FILING DATE: 2002-08-08

; PRIORITY FILING DATE: 2002-08-08

; PRIORITY FILING DATE: 2002-03-07

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: Patent in version 3.2

; SEQ ID NO 35

; LENGTH: 1375

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-10-383-930-35

Query Match

Best Local Similarity 40.0%; Score 2893.5; DB 4; Length 1375;

Matches 634; Conservative 211; Mismatches 412; Indels 175; Gaps 30;

Qy 1 MEKNLYKLVKLVKVKQVAVGVT--VTLSELAGGVVAAADNNNDGTSVQVKNKVPSPDK 58
Db 1 MEKKVRFKLVKLVKVKQVAVGVT--VTLSELAGGVVAAADNNNDGTSVQVKNKVPSPDK 49

Qy 59 FDAQAQNGQLAQAAMFKAANQAD--QTATTSQVSPATDGRVDN--QVTPAANQPAANVANQDVA 116
Db 50 --ESQASLVTTSEAAKETLTATDTSTATSATSQTAIVTDVNTSTNQSTNTTANTAN--FVV 107

Qy 117 NPA-----TDAGALNRQSAADTSTGKAVPQTS----- 145
Db 108 KPTTSEQAKTDNSDKIITTSKAVNRLTAT-----GKFPANNTAHPKVTVDKIVPIK 162

Qy 146 -----QPG-----HLETVDGKTYVDANGQRLKNYKSVIWDGTYTFPGQT 185
Db 163 KIGKLQPSLSQDDIAALGNVKNIRKVGKYYKYKEDGTLOKYNALNINGKTFPFDETG 222

Qy 186 GEAOTDLP--KTGOANODNVPDSYOANNOAYSNEASSFETVDNYLTADSWYRPRKILKNGQ 244
Db 223 ALSNNTLPKKGNIITNDNTNSEFAQYNQVYSTDVANFEHVDHYLTAEWTRPKYILKDKG 282

Qy 245 SMOASSEGDLRPLMTWPDAAATKAAAYANFWAKEGLISGSYRQNSA--NLDAATQNTQSA 302
Db 283 TWTQSTEKDFRPLMTWPDQETQRYVYVMAQLGHTQYNTATSPQLNLAAQTITQK 342

Qy 303 IEKLIASEGNTWLRDKMSQFVKSQNQSIASENETVYPNDIMQGGALLFS--NSKQTEH 361
Db 343 IEKITAETKNTWLRQTIISAFVKTSQAMNSDSEK-----PFDDHLQKGLLYSNNKLTQ 398

Qy 362 ANSDWRLNRPNTFOTGKO--KYET--TNYAGVELLLANDVDSNPNVVOBQLNHLHVLN 418
Db 399 ANSNVRLNRPNTFOTGKO--KYET--TNYAGVELLLANDVDSNPNVVOBQLNHLHVLN 458

Qy 419 WGDIVMGDKXANFDGVRVDVADNVNADLLQIQDYKAKYGTQDNEKNAIDHLSILEAWS 478
Db 459 FGNIVANDPDANFDSIRVDADVNDVADLLQIAGDYLKAAKGIHKNDAANDHLSILEAWS 518

Qy 479 GNDNDYVQDQNNFSLIDNDQSGMLKAFGYASAYRGNLSNLTAGLKRS--ANPDSDPV 537
Db 519 YNDTPLYLHDDGDNMINMDNRLSLIYSLAKPLNQRSGMNPITNSLVNRTDDNAETA 578

Qy 538 PNVVFIKRAHDSVQTRIAKIREKLGKTNADGLTNLTDLNKAFDIYNQDMNATDKVY 597
Db 579 PSYFIRAHDSVQDLIRNIIRTEI--NPNVVG--SFTTEIKKAFBIYNKDLATEKYYT 636

Qy 598 PNNLPYAWMLQNKOTVTRVYGYDMYTDNGQYMATKTPFYNAIETLLKGRIKYVAGGOA 657
Db 637 HYNLTALSYALLTNKSSVPRVYGYDMFTDQGYMAHKTINYEALETLLKARIKYVSGGOA 696

Qy 658 VSYKQDSSGILTSVRYGKANSADAGNTETRTNSGMALLINRPNFR--AYRNLTLNMG 715
Db 697 MRNQGVNSEIITSVRYGKALKATDGTDRTRTSGVAVIEGNNPSRLKASDRVVVNM 756

Qy 716 AAHKSQAYRPLLSTKDGATYLNDSVDVSRQYKTYDSQGNLSFSAELOSQVANAQVSGM 775
Db 757 AAHKNQAYRPLLSTKDGATYLNDSVDVSRQYKTYDSQGNLSFSAELOSQVANAQVSGM 815

Qy 776 IQVWVPVGAADNQDVRTSPSTQATKDGNIYHQSDALDSQVIYEGFSNFQAFQSPQYTN 835
Db 816 LGVWVPVGAADNQDVRTSPSTQATKDGNIYHQSDALDSQVIYEGFSNFQAFQSPQYTN 875

Qy 836 AVIAKNGDLFKSWGITQFEMAPQYVSSDEGTFLLDSVILNGYAFSDRYDLAMSKNKYGS 895
Db 876 VVIKKNVDFAEWGVTFDEMAPQYVSSDEGTFLLDSVILNGYAFSDRYDLAMSKNKYGS 935

Db 1116 KNAFISLGNWYFDNNGWVTGAQSIINGANYFSLNSGIQLRNAIYDNGNKVLSYYGNDG 1175
Qy 1132 RTWTKGFVEFRVDGQKWRYPNGDGTIAIGLVSLDNRTLIFYDAYGYQVKGTVTINGKS 1191
Db 1176 RRYEN-GYLLF--GQ-QWRYPO-NGIMAVGLTRVHGAVQYFDASGFQAKGQFI----- 1223
Qy 1192 YTFDADQGLVQTDNANPAPQOAGWKLLGDNQWGYRKDGQLLTGEQTIDGQKVPFDNG 1251
Db 1224 ----- 1223
Qy 1252 VQVKGGTATDASGLRFRYDRDOGHQVKGWYSTDDNWVYVNESGOVLTGLQITIDGQTVY 1311
Db 1224 -----TTADGKLYRFDPSGNQISNRFVRNSKGEWFLFDHNGVAVTGTVTFNQRLY 1275
Qy 1312 FDDKGIQAKGKAVWDENGLRYFDADSGNMLRDW-KNYVDGNWYFNRLA 1362
Db 1276 FKPNGVQAKGEFIRDANGYLRYVDPSNGNEVRNRFVRNSKGEWFLFDHNGIA 1327

Search completed: February 11, 2006, 20:57:20
Job time : 133.176 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 7.76765 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-39
Perfect score: 7230
Sequence: 1 MEKNLYKLHKVKQWVAIG.....WKNVDGNWYFNRNGLATRW 1365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New.*
1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB_PEP.*
2: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB_PEP.*
3: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB_PEP.*
4: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB_PEP.*
5: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB_PEP.*
6: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB_PEP.*
7: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB_PEP.*
8: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	367.5	5.1	2710	7 US-11-051-453-41	Sequence 41, Appl
2	334.5	4.6	2367	7 US-11-051-453-42	Sequence 42, Appl
3	290	4.0	1155	6 US-10-793-626-1780	Sequence 1780, Ap
4	280	3.9	726	7 US-11-052-554A-217	Sequence 217, App
5	270.5	3.7	690	7 US-11-052-554A-232	Sequence 232, App
6	248.5	3.4	396	7 US-11-022-562-228	Sequence 228, App
7	245.5	3.4	1992	7 US-11-013-759-3	Sequence 3, Appli
8	245.5	3.4	1992	7 US-11-013-759-13	Sequence 13, Appl
9	245.5	3.4	2047	7 US-11-013-759-4	Sequence 4, Appli
10	245.5	3.4	2047	7 US-11-013-759-7	Sequence 7, Appli
11	244.5	3.4	701	7 US-11-052-554A-231	Sequence 231, App
12	242.5	3.4	2314	7 US-11-013-759-11	Sequence 11, Appl
13	240.5	3.3	619	7 US-11-052-554A-229	Sequence 229, App
14	238.5	3.3	1290	6 US-10-485-517-141	Sequence 141, App
15	231	3.2	2053	7 US-11-013-759-9	Sequence 9, Appli
16	229.5	3.2	744	6 US-10-873-528-184	Sequence 184, App
17	228.5	3.2	332	6 US-10-873-528-190	Sequence 190, App
18	226.5	3.1	5024	6 US-10-793-626-2964	Sequence 2964, Ap
19	218.5	3.0	340	6 US-10-873-528-188	Sequence 188, App
20	216.5	3.0	1588	7 US-11-052-554A-280	Sequence 280, App
21	216	3.0	2399	7 US-11-052-554A-92	Sequence 92, Appli
22	213	2.9	658	6 US-10-873-528-17	Sequence 17, Appl
23	213	2.9	677	6 US-10-873-528-155	Sequence 155, App
24	212.5	2.9	1263	6 US-10-485-517-127	Sequence 127, App
25	212	2.9	693	6 US-10-873-528-185	Sequence 185, App

ALIGNMENTS

RESULT 1

US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-PEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: MJ1-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; CURRENT FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; PRIOR FILING DATE: 2004-09-28
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: Patentin Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 5.1%; Score 367.5; DB 7; Length 2710;
Best Local Similarity 19.4%; Pred. No. 1.2e-13;
Matches 265; Conservative 194; Mismatches 507; Indels 401; Gaps 65;
QY 240 LKNGSQWQASSSGDLRPILMTWPDAAAT-KAAYANFWAKEGLISGSYRQNSANLDAATQN 298
Db 1309 IRNKUSYFDGAGGYSLLSYPISTNINLKKDDLWI-----FNIDNEVRE 1355
QY 299 IQSAIEKTIASRGNTNWLDRKMSQFVKSQNSIASENETV-----YPNQDH----- 345
Db 1356 I--SIENGTFKKGKL--IKDVLKIDINKKLIIG--NOTIDFGSIDNKRVIYFTCEL 1409
QY 346 -----MQGALLFNSKDTTEHANSWRLLRNPTFTQKQKFTTNYAGYEL 392
Db 1410 DDKISLIIBINLVAKSYSLLLSGDK-----NYLISNLSNTIE-----KINTLGDS 1455

Sequence 191, App
Sequence 90, Appl
Sequence 238, App
Sequence 5, Appli
Sequence 8, Appli
Sequence 30, Appl
Sequence 260, App
Sequence 91, Appl
Sequence 281, App
Sequence 1, Appli
Sequence 3154, Ap
Sequence 2, Appli
Sequence 211, App
Sequence 212, App
Sequence 45, Appl
Sequence 27, Appl
Sequence 28, Appl
Sequence 25, Appl
Sequence 95, Appl

Db 1625 ICENNNIQFYFKFNLENTNYLYVGNRQNMIVPEPNYDLD-----DSGD 1669
QY 733 GIATYLDNDVDSDROYKYTDQGNLSPSASELOSANQAQSGMIQVWVPVGAADNDQDVRT 792
Db 1670 ISSTVIN-----FSQKLY-----GIDSCVNKV-----I 1694
QY 793 SPSTQATKQGNVHQSDALDSQVYEGFSNFQAFQSPDQYTNVIAKNGDLFKSWGITQ 852
Db 1695 SP-----NLY--TDEINITPVYETNNTYPEVILVDANYINEKINVIN----- 1735
QY 853 FEMAPQVVSSEDG-----TFDLSVILN--GYAFSDRYDLAMS 887
Db 1736 -DLSIRVYVNDGNDNDFLMTSTSEENKVSQVIRFVNVFKDITLANKLSFNFSQKQDVPVS 1794
QY 888 KNNKYPQDLANAIGLOAGIKVLSDLVPNQLYNLPGKEVVVATRVNQGQAKSGAT- 946
Db 1795 -----EILSFTSYEDGL-IGVDLGLVSLN-----EKFYINFGVMVSGLIY 1838
QY 947 INKTPY-----VANTRSYGDYQEGYGGKFLDDLQKLYPRLFSTKQISTGKPIDPSV 997
Db 1839 INDSLIFYFKPPVNNLITGFVTGD-----DKYFNPINGGAASIGETI---I 1882
QY 998 KITWAKYFNGSNILRGAKYVLSBGNKYNLADGKFLP--TVLNTYTGQOVANGF 1055
Db 1883 DKNY---YFNQSGVLQTVG-FSTEDGFKY-----FAPANTLDENLEGE-AIDFTGK 1929
QY 1056 ISKNGGHIYLDKNGQVNRFKISGSWYVYFSD-GKMATGKTIGNDTYLFWPNGKQLK 1114
Db 1930 LIIDENIYYPEDNYRGAV-EWKELDGMHYFSPETGKAFKGLNQIGDDKYFNSDGMQK 1988
QY 1115 EGVYDGKAYYDDNGRTWTKGFVEFRVDG-----QDKWRYF--- 1153
Db 1989 GFVSINDNK-HYFDDSG--VMKVGYTE--IDGKHFEAENGEMQIGVENTEDGKYFAH 2043
QY 1154 -----NGDGTAGLVSLDNRTLYFDAYGYOVQGTQVTTINKSYSTFDADQDLVQTON 1206
Db 2044 NEDLNGEEGEISYSILNFKNIYFDDSFYAVVGWKOLEDGSKYFYFDED----- 2094
QY 1207 ANPAPQAGWKLGNQYRGKDGOLLTGEQITIDGQVFFQDNGVQVKG----- 1256
Db 2095 ---TAEAYIGLSINDGQYFNDGGMQVGFVTTINDKVFSDGIESGVQVINDNYFY 2151
QY 1257 -----GTATDASGLR-----FYDRDQGHQVKG 1280
Db 2152 IDDNGIVQIGVPTSDGYKYFAPANTVNDNIYQAVEYSGLVRVGEDVYFYETIETG 2211
QY 1281 W---YSTSDNWNVYVNESQVLTGLQITDQTVYFDDKGIQAKGKAWDENGNLRYFDAD 1337
Db 2212 WIYDMENESDKYFVPETKACKGINLIDDIKYFDEKGMRTGLISFENN---NYFYNE 2268
QY 1338 SGNMLRDRKNDGNWYFNRL 1361
Db 2269 NGE-IQFGYINIEDKMYFEGDGV 2291

RESULT 3

US-10-793-626-1780
; Sequence 1780, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHNS
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 1155
; TYPE: PR

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1780

Query Match 4.0%; Score 290; DB 6; Length 1155;
Best Local Similarity 19.9%; Pred. NO. 1.4e-09;
Matches 287; Conservative 185; Mismatches 547; Indels 422; Gaps 66;

QY 1 MEKNLYKLHYKQKQWVAIGVITVTLSPAGGQV--AADTNNDGTSVQVNMVSDPK 58
Db 9 INKRVDFLGNKNKYSIR-KFTVGTASILVGAITLMEGAADNEAKAAEDNQLESASKEEQ 67
QY 59 FDAQQNGQLAQAMFKAANQADQATATSOVSPATD-GRVDNQVTPAANQAPAA--VANQD 114
Db 68 GSRDNESKLNQVDLNGSHSSE-KTNNVNAATEVKVEAPTSDVSKPKANEAVTNES 126
QY 115 VANPATDAGALNRQSAADTSTDGKAVPQTSQDQGHLETVGKTYVYDANGQRLK---NYS 171
Db 127 TKPKTTEAPTVEESIAET-----PKTS-----TTQDST---EKNPSSLKDNLS 170
QY 172 MVIDKTYTFDQGTGEAQTDLPKTGOANQDNVDPDSYQANNOQVSNASSFETVDNLTAD 231
Db 171 STTSKESKTDHSTKQAQMSITKNSLNDTNDSPQSEKTSQANN-----DSTDN----- 219
QY 232 SWYRPRKILKNGQSQASSEGDLRPILMTWMPDAATKAAYANFWAKEGLISGSYRQNSAN 291
Db 220 -----QSAPSKQLDSKPEQKYTKFNDPTQDVEHT-----TTKLKTPSIS 262
QY 292 LDAATQNIQSAIEKTIAS-----EGNTNLRDKMSQFVKSQ---NOMSIASENETVY 340
Db 263 TDSSVNDKQDYTRSAVASLGVDSNETEATITNAVRDNLDLKAASREQINEAIIAEALKKDF 322
QY 341 PHQDEMGG-----ALLFSNKDTEHANSDWRL-----LNRNPTFQTKQKYFTTNVAGYE 391
Db 323 SNPDY---GVDTPALNTSQSKNSPKHSASPRMNLMSLAEP---NSGKN----- 366
QY 392 LLLAND-VDNSPVPVQAEOLNHLHYLMNMGDITVMGDKDANFDGVRVDAVDNVDLQIQ 450
Db 367 ---VNDKVIITPTLSLKNSSN---HANVWPTSNEQFN----- 400
QY 451 RQYKAKYTDQNEKNAIDHLSILEAWSGNDNDYVKDQNNFSLSIDNDQSGMLKAFGYA 510
Db 401 ---LKANYELD-----DSIKEGDTFIKYQYIRPGGLEPAIK 436
QY 511 SAYRGNLSMLATAGLKNRSANPDSVPVNYFIRAHDSVQVTRIAKIIREKLGKTNADGL 570
Db 437 TQLRSKDGSIANGVYDKTNTTNTTNYVDQYQNIITGSFDLIATPKRETAIKDNQNP 496
QY 571 TWLTL--DDLKAF--DIYNQDMNAT-----DKVYVNNLPMAVAMWLQNKDVTVRVY 619
Db 497 MEVTIANEVKKDFIVDGNKDKNTTTAAVANVDV---NNKNEVVYLNQNNQNPYAK 553
QY 620 YGDMYTDNGQYMATKPPFYNAIET-----LLKGRIKYVAGQAVSYKQDWSGSL 669
Db 554 YFST-VKNGKFIPEGVKVEYVTDITNAMVDSFNPDLNSSNVKDVTSQFTPKVSAD---GTR 609
QY 670 TSVRYKKG-ANSASDAGNTETRNSGALLINRNPFRAYRNLITLNM-GAAHKSQAVRPLL 727
Db 610 VDINFARSMAKGIIVTQAVRPTGTG-----NVYTEWLTTRDGTNTNDFR--- 657
QY 728 LSTKQGIATYLDSDV---DSQYKYTDSQGNLSFSASELSQVANAQVSGMLQVMPVGA 784
Db 658 -GKSTTTVTYLNSTAGQDNPTY-----SLGDYVWLDKKNKGVDQDDDEKGLAGVYVTLKD 712
QY 785 ADNQVRTSPSTQATKQGNVHQSDALDSQVYIEGFSNFQAFQSP-DQYTNVIAKNGD 843
Db 713 SNNRELQ-----RVTTDQSGHYQFNLQNGTYTVFEAFIPDNTYTPSPANNSTDAIDSDGE 767
QY 844 LFKSWGITQFEMAPQVVSSEDGTFDLSVILNGYAFSDRY---DLAMSKNKYKSGKODLAN 900
Db 768 ---RDGTRKVVVAKGTINNADNMTVDT-----GFYLTPKYNVGVYVWEDTNKQGIQDD--- 817

```
QY 901 AIKGLQSAGIKVLSDLVNLPGKEVVTATRVNOYGOAKSGATINKTPYVANTRSYG 960
Db 818 -----NEKGISNVKVLKN-----KNGDTIGTT-----TTDSNG 846
QY 961 DYQEQYGGKFLDDLQKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 847 KYE-----FTG----- 852
QY 1021 LSEGNKYLNADGKLFPTVLNNYQOPQVSANGFISKNGIHYLDKNGQEVKNRPFKEIS 1080
Db 853 LENGDTIEPTEGTYPTKQN-----SGSDSGKDSNGTKITVTYVK--- 893
QY 1081 GSWYFDSGKMATGKTIGNDYLFMPNGKQKLEGVWDGKAYYYDDNGRTWTNKGFFV 1140
Db 894 -----DADNKT-----DSGFYKPI-YNLGDYVWEDTNKDGIQDD-----SEKGIS 933
QY 1141 EPRVDGDKWRYPNGGTTAIGLVSLDNRFLYDAYG-YQVKGQTVTINGKSYTFD- 1195
Db 934 GVKVTLUKDK-----NGN-----AIG-----TTTTDASGHYQFKG-----LENGSYTVEFETP 975
QY 1196 -----ADGDLVQTDNANPAPQG-----QAGW-----KLLGDNQW-GYRKDGQ 1232
Db 976 SGYTPTKANSQGDITVDNSGITTGTIINGADNLITDSGFYKTPKYSVGDVWEDTNKDG 1035
QY 1233 LLTGEOTIDGQKVFQDNGVQVKGTTATDASGVLRFYDRDQGHV-----GKGWYST--- 1284
Db 1036 QDDNERGISGVKVLKDEKGNIIITTTTIDENKYOQFDNLDNGNYIIHFEPKPEGMTQTAN 1095
QY 1285 --SDDNWWYVNESGQVLTGLQTDGQTVY-----PDKGIGQAKGKAVMDENGNLRYFDAD 1337
Db 1096 SGNDDE-----KADGEDVRVITTDHDFSI-----DNG-----YFDDD 1129
QY 1338 S 1338
Db 1130 S 1130
```

```
RESULT 4
US-11-052-554A-217
; Sequence 217, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 217
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-217 )
Query Match 3.9%; Score 280; DB 7; Length 726;
Best Local Similarity 34.1%; Pred. No. 2.9e-09;
Matches 62; Conservative 30; Mismatches 66; Indels 24; Gaps 5;
QY 1193 TEDADQGDILVQTDNANPAP-----OQAGWKLLGDNQWGYRKDGQLLTGEQITDQ 1243
Db 127 SFGQQAHAHVQKMTTEAKTTPSHSINTFVNDGNGWYILG-----ADGRNVTSHTIGGK 179
QY 1244 KVPFQDNGVQKGGTATDASGVLRFYDRDQGHVQKGW-----YSTSDNWWYVNESGQVLT 1300
Db 180 TMTYPAQDGKQVKGAFQAQDSGKNGHYDRDSGEN-----WTRNRFVNDQGNWYLLNNDGVPVT 235
```

```
QY 1301 GLQTDIGQTVYFDDKGIQAKGKAVMDENGNLRYFDADSGNMLDRWKNQVNDGNNYFNRRNG 1360
Db 236 GSITVNGSLYFNSDGSQVKGNFV-EEDGSLRYDKNSGDLURKTSRTINGVNYQPDNDG 294
QY 1361 LA 1362
Db 295 NA 296
RESULT 5
US-11-052-554A-232
; Sequence 232, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-07-20
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 232
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-232
```

```
Query Match 3.7%; Score 270.5; DB 7; Length 690;
Best Local Similarity 20.7%; Pred. No. 9.9e-09;
Matches 133; Conservative 102; Mismatches 220; Indels 189; Gaps 29;
QY 743 VDSRQYKYTDSQGNLSP-SASELQSVANNAQVSGMIQVWPVGAADNQDVRTSPSTQATKD 801
Db 188 IGTSAPFSQKLUKLTFTSSSKLELISHAFANL-----SNLEKLTLPKSVKTLG 237
QY 802 GNIYHQSDALDSQVIEGFSNFOA-----FAQSPQYTNNAVIAKNGDLFPKSWGITQFEMA 856
Db 238 SNLFRLTSLKHVDVEEGNESFASVDGVLFSKDKTQLIYYPSQKNDES YK----- 287
QY 857 POYVSEBDGTFDLSVILNGYAFSDRYDLAMSKNNKYGSKQDLANAIKGLQSGAKIKVLSDL 916
Db 288 -----TPKETKELASYSFN-----KNSYLKKELEL-----EGLEKIGTFAFADA 326
QY 917 VPNQLYNLPKKEVVTATRVNOYGOAKSGATINKTPYVANTRSYGDYQEQYGGKFL-----D 972
Db 327 IKLEEISLP-NSLETIERLAFYGNLE-----LKEILILPDNVKNFGRHVMNGLPKFLTLSGN 391
QY 973 DLQKLYPRLFTSKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAK-YVLSEGNKYL--- 1028
Db 382 NINSL-PSFFLSGVLDSLKEIHIKNKSTEFYVK--KDTFAIPETVKFYVTSEHIKDVLS 438
QY 1029 NLADGKFLPTVLNN-----TYGQPOVSAN-----GPI SKNGIHYLDKNGQEVKNRPFKEI 1079
Db 439 NLSTNDIIVEKVDNIKQETDVAKPKNSNQVGVWGXDKGLWYLYNESGSMATGWVKD- 497
QY 1080 SGSWYFDSGKMATGKTIGNDYLFMPNGKQKLEGVWDGKAYYYDDNGRTWTNKGFFV 1139
Db 498 KGLWYLYNESGSMATGWVK-----DKGLW-----YLYNESGSMAT--GW 534
QY 1140 VEFVRVDQDK--WRYFNGDGTIAIGLVSLDNRFLYDAYGYQVKGQTVTINGKSYTFDAD 1197
Db 535 V-----KDKGLWYLYNESGSMATGWVK-----D 557
QY 1198 QGDVLVQTDNANPAPQAGKLLGDNQWGY-RKDQGLLTGEQITDQKVFQDNGVQVKG 1256
Db 558 KG-----LWYLYNESGSMATG-----WVKDKGL----- 580
QY 1257 GTATDASGVLRFYDRDQGHVQKGWYSTSDNWWYVNESGQVLTGLQTDIGQTVYFDDKG 1316
```

Db 581 -----WYLNESGSMATGWK-DKGLWYLNESGSMATGWKDKGLWYLNESG 628
Qy 1317 IQAGKAVDENGRLRYFDADSGNMLRDRKNDGNGWYFNNG 1360
Db 629 SMATG---WYKDKLWYLNESGSMATG-WYKSGKYYTNSG 668

RESULT 6

US-11-022-562-228

; Sequence 228, Application US/11022562

; Publication No. US20050249742A1

; GENERAL INFORMATION:

; APPLICANT: Ruprecht, Ruth M.

; APPLICANT: Shisong, Jiang

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING

; FILE REFERENCE: DPN-043CN

; CURRENT APPLICATION NUMBER: US/11/022,562

; PRIORITY FILING DATE: 2004-12-22

; PRIOR APPLICATION NUMBER: PCT/US03/20322

; PRIOR FILING DATE: 2003-06-27

; PRIOR APPLICATION NUMBER: 60/392718

; PRIOR FILING DATE: 2002-06-27

; NUMBER OF SEQ ID NOS: 340

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 228

; LENGTH: 396

; TYPE: PRT

; ORGANISM: Clostridium difficile

US-11-022-562-228

Query Match 3.4%; Score 248.5; DB 7; Length 396;
Best Local Similarity 24.1%; Pred. No. 9.1e-08;
Matches 102; Conservative 67; Mismatches 162; Indels 93; Gaps 21;

Qy 969 KFLDDLOKLYPRFSTKQISTGKIDPSVKITWNAKYFNGSN-ILGRGAKYVLSGNYK 1027
Db 2 EYYPEIIVLPNPTFHKK---VNINLDSSEYKWKSTE---GSDFILVR---YLESSNKK 51

Qy 1028 LNLADGKLFPLTVLNNYGOPOVSAN-----GFISKNGGIHYLDKNGQEVKVR--- 1075
Db 52 IL---QKIRIKGLSNTKSFNKMISDFKIKLSLGIYMSN---FKSFENSELDRDHL 104

Qy 1076 -FKEISGSWYFDSCKMATGKTKIGNDTYLFMPNGKQLKEGVW--YDGKAYVYDDNGR 1132
Db 105 GFKLIIDNKYYDEASKVLKGLININNSLFYDPIDESNLVTG-WQTINGKK-YIFDIN-- 160

Qy 1133 TWTNKGFEFRVDGDKRYFNGDGTIAIGLSLDRNTLYF---DAYGYQVKGQTV--- 1185
Db 161 --TGAASTSYKII-NGKHFFVNNNGVWQLGVFKPGDGFYFAPANTQNNIEQAIVYQS 217

Qy 1186 ---TINGKSYTFDADGDLVQTDNANPAPCGQAGKWLGDNGQYR-KDQQLLTGQTID 1241
Db 218 KFLTLNGKYYFDNDS-----KAVTGMQTDGKKYFNLTAEAATGWTID 264

Qy 1242 GOKVFPDNGVGVKGTATDASGLVRFYDRDQHQVG--KG-----WYSTSD----- 1286
Db 265 GKXYFNTN-TSIASGYTIINGKHFFYNIDGIMQIGVFKGPNGFYFAPANTDANNIEG 323

Qy 1287 -----DNWVYNESQVLTGLQTDIGQTVTFDDKGIQAKGKAVMDENGRLRY 1333
Db 324 QAIRYQNRFLYLDHNIYYFGNNSKAVTGMQTINGNYYFMPDPTMAAAGLFEIDGVIYF 383

1334 FQAD 1337

384 FGVD 387

RESULT 7

US-11-013-759-3

; Sequence 3, Application US/11013759

; Publication No. US20050249747A1

; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-3

Query Match 3.4%; Score 245.5; DB 7; Length 1992;
Best Local Similarity 19.0%; Pred. No. 1.3e-06;
Matches 307; Conservative 182; Mismatches 574; Indels 549; Gaps 75;

Qy 5 LRYKLHKVKQWVAIGVTTVTLFSLAGGQVVAADTNNNDGTSVQVNMVPSDPKFDAAQ 64
Db 353 LKVKLAKTLNLTENVNTTLN-----ATTTKVGSSTTAAELLSDSLTFQPN 401

Qy 65 NQOLAQM-----FKAANOADQAT-----SQVSPATDGRVDNQVTP----- 101
Db 402 TGSQSTKTVYGVNGVKTNNAAETTAAGITRITRKIGFARDGVDKQAPYLDKKQLK 461

Qy 102 -----AANQPAANVANQDVPATDAGALNQSAADTSDGKAVPQTSQDPG 148
Db 462 VGSVAITIDGIDAGNKKISNLAKGSSANDAVTIEQL--KAAPTLNAGAGISVTPE-- 517

Qy 149 HLETVDGKTYVYDANGQRLKNTSMWIDGKTYTFDGTGEAQTDLPKTGANQNVQDSYQ 208
Db 518 --ISVDAKSGNVA-----PTYNIGV-----KTTELNSDGTSDKFS 551

Qy 209 ANNQAYSNEASFETVDNYL-----TADSWYRPRKILKNGQWASSEGDLRILMTWWP 263
Db 552 VRSGTNNLVTAEHLASVYNEVNRADSAL-----QSFTVKEEDDDANAITVAK 602

Qy 264 DAATKAAYANF-----WAKEGLISGYRQNSA-----NLDAATQNTQSAI 303
Db 603 DTTKAGAVSILKLGKNGLTVAITKDGTVTFGLSQDSGLTIGKSTLNDGLT--VKDTN 660

Qy 304 EK-KIASEG---NTNW-----LRDKMSQFVKSQWMSIASENETVYPNQDH 345
Db 661 EOIQVANGIKFTNVNGSNPGTGIANARITRDKIG-FAGSDG----AVDTNKPYLDDQK 715

Qy 346 MQGGALLFSNS-----KOTEHA-----N 363
Db 716 LQVGNVKITNTGINAGKRAITGLSPTLPSIADQSSRNIELGNTIQDKKSNAASINDILN 775

Qy 364 SDWRLLNRN-----PTFQTKQKQFTTYNAGYELLAND 397
Db 776 TGFNLKNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTSKVYVDVNVDDTTIHLGT 835

Qy 398 VDSNPVQAEQLNHLHLMWNG---DIVMGDKANFDGVRVDADVNVNADLLQIQORDYV 454
Db 836 DDKKGLGVTKYKLNKTSANGNTATNFVNVSSDEDALVNA--KDIAENLT-----LAKEIH 889

Qy 455 KAKYGTQD-----NEKNVADHLSILEAWSGNNDYV-----KQNNFSLSIDNQ 499
Db 890 TTKGTADTALQTFVTKVDENNADANAIVTGQKNNANQVNTLTILKGEN--GLNKTDK 947

Qy 500 RSGMLKAGYASAYRGNLSNLATAGLKURSANDPDPVNPYVFIHADSEVQ-----TRI 554
Db 948 NGTVTFGTINTTSLGKAGKSTLNDGGLSTK--NPTG-----SQIQVGDGVKPF 993

Qy 555 AKIIREKLGKTNADGLTNLTLDL-----NKAFD-----IYNQDMNATDKVYPPNLPWA 604

Db 994 AKVNNGVGAGIDGTTTRIDRIGTGTNGSLDKSPHLKSKGINAGK----- 1043
QY 605 YAMLMQKOTVTRVYGDMYTDN-----GQYMATKTPFYNAIETLLK----- 646
Db 1044 -----KITNIQSGETAQNSHDAVTGKIYDLKTELENKISSAKTAQNSLHEFSVA 1094
QY 647.---GRIKYVAGQAVSKQDWSGILTSVRYGKANSASDAGNTETNSGMALLINRPN 703
Db 1095 DEQGN-NFTVSNPYSSYDTSKTSDDVIT-----FAGENGITTKVNGKW-----RVG 1139
QY 704 FRAYRNLT---LNMGAHRSQAVRPLLLSTKDGIAIYVLDNSDVSQYKTSQGNLSFS 760
Db 1140 IDOTKGLTTPKLTGHN-----NNGKIV-----JDSQ-----NQONTITGL 1175
QY 761 ASELQSVANAQVSGMIQWVPVGAADNQDVRTSPSTQATKQGNIIYHQS-----ALDSQVI 816
Db 1176 SNTLANVTN-----DKGSVRT-----TEQGNIIKDEDKTRAASIVDVL 1213
QY 817 YEGFSHFOAFAQSPDQYT-----NAVIK--NGDLFKSWG-----TQFEMA 856
Db 1214 SAGF-NLQNGEAVDFVSTYDYNFADGNATTAKVTDYDTSKTSKVYVDVNVDDTTIEYK 1272
QY 857 PQVVS-----SEDTFLDSVILNGYAFSDRY-----DLAMSKNKYKSGKDLANAIGLQ 906
Db 1273 DKXGLVKTTTLTGTGANKFALSNOATGDALVKASDIVAHLNLTSGDIQTAKGASQANN 1332
QY 907 SAGIKVLSDLVNPQLNLPKEVVTATRVNQYQAKSGATINKTPYVANTRSYGDYQEQY 966
Db 1333 SAG-----YVDADGNKVIYDSTDNKYQAKNDGTVDKTEKAVAKDLVAQAQTPD 1381
QY 967 GKFLDLDKLYLPLRFLSTKQISTGKIDPSVKLTNWSAKYFNKSNILGRGAKVYLSEGNK 1026
Db 1382 G-----TLAQM-----NVKSVINKEQVNDANK-----KQGINEDNAFVKGLEKAASD-NK 1425
QY 1027 YLMADGKFLPTV--LNNTYGOPOQVSA--NGFISKNGGHHYLDKNGQEVKRFKE--- 1078
Db 1426 TKMAA-----VTGDLNAVAQTPFLFAGDTGTTAKKGLTILKGGQDTNKLTONNIG 1479
QY 1079 -ISGSWYFYFSDG---KMATGKTKIGNDTYLFMPNGKQLKEGVYDYGKAYYYDDNGRTW 1134
Db 1480 VWAG-----TDGFTVKLAKDLTNLSVN-----AGGTKIDDKGVSFVDSGQAK 1523
QY 1135 TNKGFEFRVDGQKRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTVTINGKSYTF 1194
Db 1524 AN-----TPVLSANGLDLGGKVISNVRKG-TK 1549
QY 1195 DADQGLVQTDNA-NPAPOQAGKLLGDNQWGYR-----KDGQ 1232
Db 1550 DTDAAVQQLNEVRNLLGLNAG-----NDNADGNQVNIADIKKDPNSGSSSRTVTIKAGT 1605
QY 1233 LITGEQITDGQKVFQDNGVQV---RGGTAT--DASGLVRFYDRDQGHQVKGWYSTSDDN 1288
Db 1606 VLKGKGNNDTEKL--ATGGIQVGVDKDGNANGDLSNV-----VWTKQKDG 1648
QY 1289 WTVVNESGVLTLGTLIDGTQVTFDDKGIQAKGKAVDENGNIURYFDADSGN 1340
Db 1649 -----SKALLATYNAAGQTNLYTNPAEADRI--NEQG-IRFFHVNDGN 1691

RESULT 8

US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759

; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1992
; TYPE: PRF
; ORGANISM: Moraxella catarrhalis
US-11-013-759-13

Query Match 3.4%; Score 245.5; DB 7; Length 1992;
Best Local Similarity 19.0%; Pred. No. 1.3e-06;
Matches 307; Conservative 182; Mismatch 574; Indels 549; Gaps 75;

QY 5 LRYKLHKVKQWVAIGVTTVTLSFLAGGVAAADTNNDGTSTVQVNMKVPSPKPFDAQAQ 64
Db 353 LKVKLAKTLNLTENVTTTLN-----ATTTVKVGSSSSTTAELLSDSLTFQPN 401
QY 65 NQLAQAM-----FKAAQADQAT-----SQVSPATDGRVDNQVTP----- 101
Db 402 TGSQSTSKTVYGVGVKFTNNAETTAAGTTRITRDKIGFARDGDVDEKQAPYLDKKQK 461
QY 102 -----AAQAPAAVNAQDVANPATDAGALNRQSAADTSTDCGKAVPQTSQDQK 148
Db 462 VCSVAITIDNGIDAGNKISNLAKSSANDAVTIEQL--KAAKPTLNACAGISVTFPE-- 517
QY 149 HLETVDGKTYVYDANGQRLKNYSMVIDGKTYVFDGQGEAQTDLPKTGOANQDNVPSYQ 208
Db 518 --ISVDAKSGNVTA-----PTYNIGV-----KTTLENSDGTSDKFS 551
QY 209 ANQAYSNEASSFETVDNYL-----TADSWYRPRKILKNGQSQWQASSEGDLRPLMTWTP 263
Db 552 VKSGTGNLSLVTAETHASLYNEVNRDADSAL-----QSFTVKEEDDDANAITVAK 602
QY 264 DAATKAAVANF-----WAKEGLISGSYRQNSA-----NLDAATQNIQSAI 303
Db 603 DTTKAGAVSILKLGKNGLTVAATKDGIVTFGLSQDSGLTIGKSTLNNDGLT--VKDTN 660
QY 304 EK-KIASEG-----NTNW-----LRDKMSQFVKSQNWSIASENETVYPNQDH 345
Db 661 EQIQVANGIKFTNVNGSNPGTGIANTARIITRDKIG-FAGSDG----AVDTNKPYLDDQK 715
QY 346 MOGGALLFSNS-----KDETEHA-----N 363
Db 716 LQVGNVKITNTGINAGGKAITGLSPTLPSIADQSSRNIELGNTIQDKKSNAASINDILN 775
QY 364 SDWRLLRN-----PTFOTGKQKPTTVYAGVELLLAND 397
Db 776 TGFNLKNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTSKVYDVNVDDTTIHLTGT 835
QY 398 VDNSNPVVOAEQLNHLHYLMNWG---DIVMGDKANFDGVRVDADVNVNADLIQIQDY 454
Db 836 DDNKKLGVKTKLNTKSANGNTATNFVNSSDDALVNA--KDIENLAT----LAKEIH 889
QY 455 KAKYGTQD-----NEKNAIDHLSILEAWSGNDNDYV-----KQNNFSLSDNDQ 499
Db 890 TTKGTADTALQFTTVKVKVDENNADANAITYGQKNNANQVNTLTLLKGEN--GLNIKTDK 947
QY 500 RSGMLKAFGYASAYRGNLSNLATAGLKNRSANPDPVPVYFIRAHDSVQ-----TRI 554
Db 948 NGTVTFGINTTSLGKAGKSTLNDGSLIK--NPTG-----SEQIQVGADGVKF 993
QY 555 AKIIREKLGKTNADGLTNLTLDL-----NKAFD-----IYNQDMNATDKVYPNNLPA 604
Db 994 AKVNNNGVVGAGIDGTTRITRDEIGTGTNGSLDKSPHLKSKGINAGK----- 1043
QY 605 YAWMLQNKDVTVRVYGDMYTDN-----GQYMATKTPFYNAIETLLK----- 646
Db 1044 -----KITNIQSGETAQNSHDAVTGKTIYDLKTELENKISSAKTAQNSLHEFSVA 1094
QY 647 ---GRIKYVAGQAVSKQDWSGILTSVRYGKANSASDAGNTETNSGMALLINRPN 703

Db 1095 DEQGN-NFTVSNPYSSYDTSKTSVDIT-----FAGENGITTKVKNQGV-----RVG 1139

QY 704 FRAYRNLT---LNMGAHKSQAYRPLLLSTKDGIAATYLNDSVDSDSQYKYTDSQGNLSFS 760

Db 1140 IDQTKGLTTPKLTGVN-----NNGKGV-----IDSQ-----NQGNITGL 1175

QY 761 ASELOSANAOVSGMIVVWVPGAADNQDVRTSPSTQATKGNIIHQSD-----ALDSQVI 816

Db 1176 SNTLANVTN-----DKGSVRT-----TEQGNIIKDEKTRAASIVDVL 1213

QY 817 YEGFSNFQAPQSPDOVT-----NAVIK-----NGDLFKSWG-----TQFEMA 856

Db 1214 SAGF-NLQNGEAVDFVSTYDYNFADGNATTAKYVDDTSKTSKVYDVVNDVTIEVK 1272

QY 857 PQYVS-----SEDTGLSDVILNGYAFSDRY-----DLAMSKNKYKSGKDLANAKGLQ 906

Db 1273 DKKLGVKTTLTSTGTGANKFALSNOATGALVKASDIVAHLNTLSGDIQTAKGASQANN 1332

QY 907 SAGIKVLSLDVFNQNLNLPKQEVVTAATRVNQYQAKSGATINTPTVANTRSYGDYQEQY 966

Db 1333 SAG-----YVDADGNKVIYDSTDNKYQAKNDGTVDKTEVAKDKLVAQAQTPD 1381

QY 967 GKGFLDLOKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSGKNK 1026

Db 1382 G-----TLAQM-----NVKSVINKEQVNDANK-----KQINEDNAPFVKGLEKASD-NK 1425

QY 1027 YLNADGKFLPLPTV---LNNYTGPOQVSA---NGFISKNGGIHYLDKNGQEVKGRFKE----- 1078

Db 1426 TKNA-----VTGDLNAVAQTPFLTFAGDTGTAKKLGTLTKGQDITNKLTDNNIG 1479

QY 1079 -ISGSWYFSDG---KMATGKTKIGNDYFLFMPNGKQKKEGWYDGKAYYDDNGRTW 1134

Db 1480 VWAG-----TDGFTVKLAKDLTNLSVN-----AGGPKIDDKGVSFVDSGQAK 1523

QY 1135 TNKGFVEFRVGDQKWRYFNGDGTIAIGLVSLDNRTLYFDAYGVYQKQTVTLNGSYTF 1194

Db 1524 AN-----TPVLSANGLDGGKVISNVKG-TK 1549

QY 1195 DADQGLVTDNA-NPAQOQAGKLLGDNQMYR-----KDGQ 1232

Db 1550 DTDAAVQQLNEVRNLLGLNAG-----NDNADGNQVNIADIKDPNSGSSNRTVIKAT 1605

QY 1233 LTGBOITDQKVFPODNGVQV---KGGTAT-DASGVLFYRDRDQGHQVKGWYSTSDN 1288

Db 1606 VLGGKGNNDTEKL--ATGIGIQGVDDKQGNANGDLSNV-----WVKQKQK 1648

QY 1289 WYVNESGVLTGLQITDQTYFDDKGIQAKGKAWDENGNLRYPDADSGN 1340

Db 1649 -----SKKALLATYNAAGOTNYLTNNPABEADRI--NEQG-IRFPFVNDGN 1691

RESULT 9

US-11-013-759-4

; Sequence 4, Application US/11013759

; Publication NO. UB20050249747A1

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Sasaki, Ken

; APPLICANT: Yang, Yan Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE

; TITLE OF INVENTION: PROTEIN OF MORAXELLA

; FILE REFERENCE: 1038-921MIS:jb

; CURRENT APPLICATION NUMBER: US/11/013,759

; CURRENT FILING DATE: 2004-12-16

; PRIOR APPLICATION NUMBER: US/09/361,619

; PRIOR FILING DATE: 1999-07-27

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 4

; LENGTH: 2047

; TYPE: PR1

; ORGANISM: Moraxella catarrhalis

US-11-013-759-4

Query Match 3.4%; Score 245.5; DB 7; Length 2047;
Best Local Similarity 19.0%; Pred. No. 1.4e-06;
Matches 307; Conservative 182; Mismatches 574; Indels 549; Gaps 75;

QY 5 LRYLHKVKKQWVAIGVTTVTLISFLAGQGVAAADTNNNDGTSVQVKNQWPSDPKFDQAQ 64

Db 408 LKVKLAKTLNLTENVNTTLN-----ATTTVKVGSSSTTAELLSLSLFTFQPN 456

QY 65 NGOLAQAM-----EKAANOQDOTAT-----SQVSPATGRVDNQVTP----- 101

Db 457 TSQSSTSKTVYGVNGVKFTNNAETTAAGTTRIDRKIGFARDGVDKQAPYLDKKQLK 516

QY 102 -----AANQPAANVANQDVANPATDAGALNRQSAADTSDGKAVPQTSQDPQ 148

Db 517 VGSVAITIDNGIDAGNKKISNLAKGSSANDAVTIEQL--KAAKPTLNAGAGISVTPTE-- 572

QY 149 HLETVDGKTYVVDANGQRLKNYSMVJDGKTYYPDGTGEAQTDLPTKGTQANQDNVPSYQ 208

Db 573 --ISVDKSGNVA-----PTYNIGV-----KTTLELNSDGTSDKFS 606

QY 209 ANNQAYSNEASSPFTVDNYL-----TADSWYRPRKILKNGQSQWASSEGDLRILMTWP 263

Db 607 VKSGTNNSLVTAELASLYLNEVNRADSAL-----QSFTVKEEDDDANAITVAK 657

QY 264 DAATKAAVANF-----WAKEGLISGYSRONSA-----NLDAATONIQSAI 303

Db 658 DTTKAGAVSILKKGKGLTVATKDGVTYFGLSQDSGLTIGKSTLNDGLT--VKDTN 715

QY 304 EK-KIASEG-----NTNW-----LRDKMSQFVKSQNQSASISASNETVYPNQDH 345

Db 716 EQIQVANGIKFTNVNGSNPGTGIANTARITRDKIG-FAGSDG---AVDTNKPYLDQDK 770

QY 346 MOGGALLFNS-----KDEHA-----N 363

Db 771 LQVGNVKITNTGINAGKKAITGLSPLPSIADQSSRNIELGNTIQDKKSNAASINDILN 830

QY 364 SDWRLLRN-----PTFQTQKQKFTTVNAVGYELLAND 397

Db 831 TGFNLKNNNPIDFVSTYDIVDFANGNATATVTHDTANKTSKVYDVVNDVTIHLTGT 890

QY 398 VDNSNPVVOAEOQLNHLHYLMNWG---DIVMGDKDANFDGVRVDADVNNADLLQIQDY 454

Db 891 DDNKKLGKVKTKLNTKSANGNTATNFVNSSDEDALVNA--KDIAENLAT---LAKEIH 944

QY 455 KAKYGTQ-----NEKNAIDHLSILEAWSGNDNDYV-----KQNNFSLSDNDQ 499

Db 945 TTKGTADTALQTFYVKVVDENNADANAITYGQKNANNQVNTLTILKGEN--GLNIKTDK 1002

QY 500 RSGMLKAFGYASAYRGNLSNLATAGLKNRSANPDSPPVNYVYFIRAHDSEVQ-----TRI 554

Db 1003 NGTVTFGINTTSGLKAGKSTLNDGGLSIK--NPTG-----SEQIQVAGGVKF 1048

QY 555 AKIIREKLGKTNADGLTNLTLDL-----NKAFD-----IYNQDMNATDKVYPPNLPMA 604

Db 1049 AKVNNVGVGAGIDGTTRITRDEIGFTGNGSLDKSPHLKSDGINAGK----- 1098

QY 605 YAMMLQNKDVTVRVYGYDMYTN-----GQVMAKTTPYNAIETLLK----- 646

Db 1099 -----KITNIQSGEIAQNSHDAVTGGKIYDLKTELENKISSAKTAQNSLHEFSVA 1149

QY 647 ---GRIKVVAGQAVYKQDWSGILTYSRYKGSANSADAGNCTETNSGMALLINRPN 703

Db 1150 DEQGN-NFTVSNPYSSYDTSKTSVDIT-----FAGENGITTKVKNQGV-----RVG 1194

QY 704 FRAYRNLT---LNMGAHKSQAYRPLLLSTKDGIAATYLNDSVDSDSQYKYTDSQGNLSFS 760

Db 1195 IDQTKGLTTPKLTGVN-----NNGKGV-----IDSQ-----NQGNITGL 1230

QY 761 ASELOSANAOVSGMIVVWVPGAADNQDVRTSPSTQATKGNIIHQSD-----ALDSQVI 816

Db 1231 SNTLANVTN-----DKGSVRT-----TEQGNIIKDEKTRAASIVDVL 1268

1
QY 817 YGFSNFQAFAPSPDOYT-----NAVIAK--NGDLFKSWG1-----TOFEMA 856
DB 1269 SAGF-NLQNGEAVDFVSTYDTVNFADGNATTAKVTDYDTSKTSKVYDVNVDDTTIEVK 1327
QY 857 PQYVS-----SEDGTFDLSVILNGYAFSDRY-----DLAWSKNKYGSKODLANA1KGLQ 906
DB 1328 DKKLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIVAHNLTLSGDIQTAKGASQANN 1387
QY 907 SAGIKVLSDLVFNQLNLPKEVVTATRVNOYQCAKSGATINKTPYVANTRSYGDYQEQY 966
DB 1388 SAG-----YVDAGNKVIYDSTDNKYQAKNDGTVDKTEVAKDLVAQAQTPD 1436
QY 967 GKGFLDOLKLYRPFSTKQISTGKPIDPSVKITWSAKYFNGSNILGRGAKYVISEGNK 1026
DB 1437 G-----TLAQM-----NVKSVINKEQVNDANK-----KQGINEDNAFVKGLEKAASD-NK 1480
QY 1027 YNLADGKFLPTV--LANNYGOPOVSA--NGFISKNGGIHYLDKNGQEVKNRFE---- 1078
DB 1481 TRNAA-----VTVGDLNAVAQTFPLTFAGDTGTAKKLGELTITIKGGQDTRNKLTDNNIG 1534
QY 1079 -TSGWYFDSQD---KMATGKTKIGNDTYLFMPNGKQLKEGVYDGGKAYYYDNGRTW 1134
DB 1535 VVAG-----TDGFTVKLAKDLTNLSVN-----AGGTKIDDKGVSFVDSGQAK 1578
QY 1135 TNKGFEFRVDGQKWRYFNGDGTIAIGLVSLDNRTLYFDAYGYQVKGQTVTINGKSYTF 1194
DB 1579 AN-----TPVLSANGLDLGGKVISNVGKG-TK 1604
QY 1195 DADQGLVQTDNA-NPAPOGAQWKLGNQWYR-----KDQO 1232
DB 1605 DTDAAVNVQQLNEVRNLLGLGNAG-----NDNADGNQVNIADIKKOPNSGSSNRVTIKAGT 1660
QY 1233 LTGQETIDQKVFQDNGVQV---KGGTAT-DASGLRFRYDRDQGHQVKGWYSTSDN 1288
DB 1661 VLGGKGNNTDKL---ATGGLQVGDQGNANGDLSNV-----WVKTKDQ 1703
QY 1289 WYVNESGQVLTLQOTIDQTVYFDDKGIQAKGKAVMDENGLRYPDASGN 1340
DB 1704 -----SKKALLATYNAAGQTNLTNNPAEADRI--NEQG-IRFFHVNDGN 1746

RESULT 10

US-11-013-759-7
; Sequence 7, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-92LMIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; PRIOR FILING DATE: 2004-12-16
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-7

Query Match 3.4%; Score 245.5; DB 7; Length 2047;
Best Local Similarity 19.0%; Pred. No. 1.4e-06;
Matches 307; Conservative 182; Mismatches 574; Indels 549; Gaps 75;

QY 5 LRYKLHKVKKQWVAIGVTTVTLSTFLAGGVVAADTTNNDDGTSYQVKNWVPSDKPDAQAQ 64
DB 408 LKVKLAKTLNLTENVNTTLN-----ATTTVKVGSSSTTAEILLSDLSLTQPN 456

QY 65 NQLQAQAM-----FKAANQADQAT-----SQVSPATDGRVNDQVTP----- 101
DB 457 TGSQSTSKTVGVNGVKFTNNAETTAA1GTTTRITDK1GFARDGDVDBKQAVLQKKQLK 516
QY 102 -----AANQPAANVANQDVANPATDAGALNRQSAADTSTDGKAVPQTSDDQ 148
DB 517 VGSVAITTDNGIDAGKKISNLAKGSSANDAVTIEQL--KAAPTLNAGAGISVTPTE-- 572
QY 149 HLETVDGKTYVYDANGQRLKNYSMWIDGKTYTFDQGTQCEAQTDLPLTQGANQNPVDSYQ 208
DB 573 --ISVDAKSGNVA-----PTYNIGV-----KTELNSDGTSDKFS 606
QY 209 ANQAYSNEASFEFVDNVL-----TADSWYRPRIKLNKQSQWQASSEGDLRPLMTWWP 263
DB 607 VKSGTNNSLVTAELASYLEVNRNADSAL-----QSFTVKEEDDDANAITVAK 657
QY 264 DAATAKAAVANF-----WAKEGLISGSYRQNSA-----NLDAATQNIQSAI 303
DB 658 DTTKNAGAVSILKLGKNGLTVATKDGTVTFGLSQDSGLTICKSTLNNDGLT--VKDTN 715
QY 304 EK-KIASEG---NTNW-----LRDKMSQFVKSQNWSIASENETVYPNQH 345
DB 716 EQIQVANGIKFTNVNGSNPGTGIANTARITRDKIG-FAGSDG-----AVDTNKPYLDQDK 770
QY 346 MOGGALLFNS-----LDRKMSQFVKSQNWSIASENETVYPNQH 363
DB 771 LOVGNVK1TNTGINAGKKAITGLSPTLPSIADQSSRNIELGNTIQDKKSNAAASINDILN 830
QY 364 SDMRLLNRN-----PTFQTKQKQFTTNYAGVEILLAND 397
DB 831 TGFNLKNNNPIDFVSTYD1VDFANGNATTATVTHDTANKTSKVYDVNVDDTTIHLTGT 890
QY 398 VDNSNPVQAEOLNHLHYLMWNG---DIVMGDKDANFDGVRVDAVDNVMNADLLQIQDYY 454
DB 891 DDNKKLGKVTTKLNTSANGNTATNFVNSSDEALVNA--KDIAENLT---LAKEIH 944
QY 455 KAKYGTQD-----NEKNAIDHLSILEAWSNDNDYV-----KQNNFSLSDNDQ 499
DB 945 TTKGTADTALQTFVTKVDENNADDANA1TVGQKNNANQVNTLT1KGEN--GLN1KTDK 1002
QY 500 RSGMLKAFGYASAYRGNLSNLATAGLKNRSANPDSDPVNPFVIRAHDEVO-----TRI 554
DB 1003 NGTVTFGINTTSLGKAGKSTLNDGSLIK--NPTG-----SEQIQVGADGVKF 1048
QY 555 AKI1REKLGKTNADGLTNLTLDDL-----NKAFD-----IYNQDMNATDKVYPPNLPWA 604
DB 1049 AKVNNNGVVGAGIDGTTRITRDEIGFTGTNGSLDKSKPHLSKDGINAGGK----- 1098
QY 605 YAWMLQNKDVTVRVYGDMYTN-----GQWMAKTFPYNAIETLLK----- 646
DB 1099 -----KITN1OSGE1AQSHDAVTTGGKTYDLKTELENKISSTAKTAQNSLHFEFSA 1149
QY 647 ---GR1KVAGGQAVSYKQDWSSGILTSVRYKGNASASDAGNTEPTRNSGMALLINRPN 703
DB 1150 DEQGN-NFTVSNPYSSYDTSKTSDDVT-----FAGENGITTKVNGKV-----RVG 1194
QY 704 FRAYRNLT---LNMGAHKSQARYRPLLSTKDG1ATYLNDSVDVSRQYKYDTSQGNLSPS 760
DB 1195 IDQTKGLTTPKLTJVGK-----NNGKGV-----IDSQ-----NQNTITGL 1230
QY 761 ASELOSVAQAQVSGMIQWVVPVGAADNQDVRTSPSTQATKDN1YHQSD-----ALDSQV1 816
DB 1231 SNTLANVTN-----DKGSVRT-----TEQGN1IKDEDKTRAAS1VDVL 1268
QY 817 YGFSNFQAFAPSPDOYT-----NAVIAK--NGDLFKSWG1-----TOFEMA 856
DB 1269 SAGF-NLQNGEAVDFVSTYDTVNFADGNATTAKVTDYDTSKTSKVYDVNVDDTTIEVK 1327
QY 857 PQYVS-----SEDGTFDLSVILNGYAFSDRY-----DLAWSKNKYGSKODLANA1KGLQ 906
DB 1328 DKKLGVKTTTLTSTGTGANKFALSNOATGDALVKASDIVAHNLTLSGDIQTAKGASQANN 1387

Db 434 TA-----EIQSGGLTFT-----PTTNASTDKTVYGTGDLKFTDINSNTALEDTTRITKXIGF 485
QY 385 TNYAGYELLANDVNSNPVQAEQUNHLHYLMNWG-----420
Db 486 SNKAG-----TVDENKPYLDKDKLVGNSTLNNGLTVNNTIGGSNKQIQVGADGKIF 538
QY 421 -----DVMGDKDANFD-----GVRV-----436
Db 539 ADVNVNVSNAKFGTTRITEEIEGFADADGKVKSPYLDKQLOVGGVKIKTKDSGINAG 598
QY 437 -----DAVNUNA-----DLOIORD-----454
Db 599 DOKISNVKDATDDTDAVTVYKQLKQVQDADGALQSPSIRDEKGOEFTINLNGWTPWT 658
QY 455 -----KAKYGTDO-----NEKNAIDHLSILEAWSGND 481
Db 659 FETITFAGENGISINDIAKGVKVGIDPINGLITTPKLTGSDKDGKTQLVIEQVASGND 718
QY 482 NDYV-J-----KOQNFSLSDNDQSRGML-----KATG 508
Db 719 TKNIIRGLSPTLPSITNAGGVRTTEQGNITTSDEDSKAASIGDILNTGFNLKNNNSVG 778
QY 509 YASAYR-----GNLSNLATAGLKNRSANPDSPVENVYFIRAHDSQVOTRIAKIIREKL 562
Db 779 FVSTYNTVPIDGNAT---TAKVYDETQTSKV-----TYDVNVDKTIETLGDN- 826
QY 563 GKTNADGLNLTLDLNL---KAPDIYNQD-----MNATDKVYYPNNLPMAYAWMLQKQTV 615
Db 827 GKTKNGKGVKTTTLTTNANGKATNPSTTDNDALVNAKD-----IAENLNTL 872
QY 616 TRYVYGDMDYDNG-----QYMATKTPFFNAIETLLKGRIKYVAGQAVSYQDWSGSL 669
Db 873 AK-----EIHTTKGTADTALQFKVKKGATDDETTIVG-----KDGTONGKT 915
QY 670 TSVRYCKGANSADAGNTETRNSGSMALLNNRPNFPAYRNLTLM- GAAHKSQAYRPLLL 728
Db 916 VNTLKLKGBGLTVATN---KDGTVFTGINTOSGLKAGDSTTLNKDGLSKINPASNEQIQ 972
QY 729 STKDGATYVND-----SDVDSRQYKYVTDQGNLSFSASELQ-----SVANAQVS 773
Db 973 VGADGVKFAKVDKGNSTGIDGTSRITKQIGFTGANGSLDTTKPHLTKDKLKVGEVET 1032
QY 774 GMIQVWVPVGAADNQDVRTSPQATKD---GNIVHQS DALDSQV---IYEGFSNFQAF 826
Db 1033 ---NTGINAGGKKITNIQSGDITQNSNDAVTGGRVYDLKTELESKINSAAKTAQNSLHEF 1089
QY 827 AQSPDQ-----YTNAVIANKGLF-----KSWGITQPEM 855
Db 1090 SVADEQGNHFTVSNPYSSYDTSKTSDVITTFAGENGITTKVKNQVVRVIGIDQTKGLTTPKL 1149
QY 856 APO-----YVSSEDG---TFLDSVILN-----GYAFSDRYDLAMSKNNKYGSKQDL 898
Db 1150 TVGNNGKGIVIDSKOGQNTITGLSNTLANVTNDGAGHALS-----QGL 1193
QY 899 ANAIKGLQAGIKVLSDLPVNQLYNLPKG- EVVTATRVNQYGOA---KSGATINTKTPYA 954
Db 1194 ANDTDTRAASI---GDVL- NAGFNLQNGEAVDF- VSTYDTPDFIDGNATTAKVTY-- 1245
QY 955 NTRSYGDYQEOYGGKFLDDLQKLYPRLF-----STKQJSTGKPIDPSVKITNWSAKYFN 1008
Db 1246 -----DOTSKTSKVYDVNVNVDNKTIETVTSKKL---GVKTTTLTKTSAN 1286
QY 1009 GSNILGRGAKYVLSEGNKYNLINADGKLFPTVLNNTYQOPQVSANGFISKNNGIHYLDKN 1068
Db 1287 GN-----ATKPSAADGALVKASD-----IATHLNTLAGDIQ-TAKGASQASSASVYDAD 1336
QY 1069 GOEVKRNRFKEISGSWYTFSDGKQKATGKTKIGNDTYLFMPNGKQKLEGVWYDOKKAYYD 1128
Db 1337 GNRV-----IYDS-----TDKTYQVN 1353
QY 1129 DNGRTWTKGFVEFRVDGQDKWRYFNGDGTIA-IGLVSLDNRTLYPDAYGYQ----- 1179

Db 1354 DKGOYDKNKEVAKDKLVAQAQ-----TPDGTTLAQMNVKSVINKEQYNDANKKQGINEDNAP 1409
QY 1180 VKG-QTVTINGSYTTFDADQDGLVQTDNANPAPQOAGWKLLGDNQMGVYRKDGQLLT--G 1236
Db 1410 IKGLENAAKDKTKKNAAVTVGDL-----NA-----VAQTPLTFAGDTGTTAKKLGETLTIKG 1461
QY 1237 EOTIDGQKVFPODNGOVQVKGGTATDASGVLRFYDRDQGHQVGKGYWSTSDDNWVYVNBESG 1296
Db 1462 GGT-DTNKL--TDNNIGVVAG--TDGFTVKLAKDJUTNLNSVNAVGGTRIDKEKISFVDANG 1516
QY 1297 QVLTLGLQITIDGTVTFDDKIGIOAKGKAVMDENG-NLRYFPD-----ADSGNML- 1342
Db 1517 QAKANTPVLSANGLDLGGKRIISNIGAAVDDNDVAVFKQFNEVAKTVNNLNQNSGASLP 1576
QY 1343 -----RDRWKNVDGNWYFNRNGL 1361
Db 1577 FVVTDANGKPINGTQDKPKQAIKGADGKYTHANANGV 1613

RESULT 13
US-11-052-554A-229
; Sequence 229, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 229
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-229

Query Match 3.3%; Score 240.5; DB 7; Length 619;
Best Local Similarity 23.6%; Pred. No. 5.1e-07; Indels 91; Gaps 12;
Matches 70; Conservative 42; Mismatches 93

QY 1066 DKNGQEVKRNRFKEISGSWYTFSDGKMATGKTKIGNDTYLFMPNGKQKLEGVWYDOKKAY 1125
Db 394 EKPA PAPKTGWQENGWYFYNTDGSMTG-----WLQN-----NGSW-----Y 432
QY 1126 YYDDNGRTWTKGFVEFRVDGQDKWRYFNGDGTIAIGLVSLDNRTLYPDAYGYQVKGQTV 1185
Db 433 YLNSNGAMAT--GWLQY-----NGSWYLYNANGAMATGWAKVNGSWTYLNA----- 476
QY 1186 TINGSYTFDADQDGLVQTDNANPAPQOAGWKLLGDNQMGY-RKDQGLLTGQTIDGQK 1244
Db 477 --NGAMAT-----GW-LQYNGSWYLYNANGAMATGWAKVNGSW 511
QY 1245 VFFQDNGVQVGKGTATDASGVLRFYDRDQGHQVGKGYWSTSDDNWVYVNBESGQVLTGLQT 1304
Db 512 YLYNANGAM-----ATGWLYQY-----NGSWYLYNANGAMATGWAK 546
QY 1305 IDGQTVYFPDKGIOAKGKAVMDENGNLRYFDADSGNMLRDRWKNVDGNWYFNRNG 1360
Db 547 VNGSWYLYNANGAMATG---WVKDGTWYLYEASGAKASQWFKVSDRKWYVYVNLGL 599

RESULT 14
US-10-485-517-141
; Sequence 141, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield

```
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629W0
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 141
; LENGTH: 1290
; TYPE: PRP
; ORGANISM: Staphylococcus aureus
; US-10-485-517-141

Query Match      3.3%; Score 238.5; DB 6; Length 1290;
Best Local Similarity 20.0%; Pred. No. 1.9e-06;
Matches 251; Conservative 155; Mismatches 481; Indels 367; Gaps 59;

QY 40 NNNDGTSVQVKNMVPSPDKFDAQONGQALQAQAFKAAQADQATATQVSPATDGRVDNQV 99
Db 187 NNVDAAAMDKLRQSI-ADNATTQKNQ-----YTDASQNKQDAYNNAVTTAQQGIIDQT 238
QY 100 TPAANQPAANVANQDVANPATDAGALN-----RQSAADTSTDGKAVPQT-SDQ 146
Db 239 SPTLD---PTVINOAGQVSTTKNALNGNENLEAAKQQAQSQSLGSLDNLNNAQKQVTTQ 295
QY 147 PGHLETVGKTYVDVANGQRLKNYSMVIDGKTYVFDGTEGAEATDLPKTKQANQDNVPS 206
Db 296 INGAHTVD-EAQQIKQNAQNL-NTAM-----GNLQQAAD-KDQATATVNTFDA 341
QY 207 YOANNOAYSNEASSPTVDNYLTADSWYPRKILKNGQSQWQASSEGLRPLMTWPPDAA 266
Db 342 DQAKQQAQVNTAVTNAENIISK-----ANGGNATQAEVEQAIRQV-----NAA 383
QY 267 TKAAYANF---WAKE---GLISGYRQNSANLDAATQNIQSAIEKKIASSEGTNNLWLRDKM 320
Db 384 KQALNGNANVQAHAKDENTALINSSNDLNQAKDALKQQQVQAT----- 426
QY 321 SQFVKSQOWMSIASENETYPNQDHMOGGALLFSNSKUTEHANSWRLLNRPFTFQTGKQ 380
Db 427 -----TVAGVNVNQTQALNNAMTQLKQGIADKEQTKADGNFYNADP-----DKQ 472
QY 381 KYFTTYAGYELLANDVD-----NSNPVQAEQNLHLHYLMNWDIVMGDKDAN 430
Db 473 NAYNOQAVAKAEALISATPDVVVTPSPBITAALNKVTOAK--NDLN-----GNTNLATAKQ 525
QY 431 FDGVRVDAVDVNDLQIQDYKAKYCTDQNEKNAIDHLSILEAWSGNDND----- 483
Db 526 VOHA-IDQLPNLN-----QAKRD-----EYSKQITQATLVNPNVNAIQQAATTLNDAMTQLKQ 576
QY 484 -----YVKDQNFSLSDNDQSGMLKAFGYASAYRGNLSNLATAG--LKNRSANPQSD 535
Db 577 GIANKAQIKGSNYH-DADTDQ-----TAY-----DNAYTKABELLQTTNPTMD 621
QY 536 PVPNTVFTRAHSEVQTRIAKII-----REKLGKTNADGLTNL-TLDDLNKAPD--I 584
Db 622 P-----NTIQQAQTKVNDTNOALNGNQKLADAKQDAKTTGLTGLDHLNDAKQAL 670
QY 585 YQDNMNATDKVYYPNPLPAYAWMLQNKDVTTRVYGDYMTDNGQYMATKTFPYNAIETL 644
Db 671 TTQVEQADP-IATVNVK-----QNAQL-----NNAMTLLNALQDKTETLINSINT 717
QY 645 LKGRITYAGQAVSYKQDWSSGILTSVRYGKGANSASDAGNETRNSGMAL-----LIN 699
Db 718 DADQAKDKDAYTNAVSH-----AEGILSKA-----NGSNASQTEVEQAMQVNEAKQALN 766
QY 700 NRPNFRVNTLNLNM--GAHKSQAYRPLLSTKDGIAITYLNDSDVDVSRQYKTYDSQGNL 757
```

```
767 GNDNVORAKDAKQVITNANDLNQAMTQL-----KQGIA-----DKD-----QTKANGNF 811
QY 758 SFSASELSQSVANAQVSGMIQV--WVPVGAADNQ-----DVRTSPSTQATKD 801
Db 812 VNADTKQAYNNNAVAHAQEIISGTFNANVDPQVQAALQOVNQAKGDLNGNHNLQVAKD 871
QY 802 G-----NIYH-QSDALDSQVIY-EGFSNFQAFQSPDQYTNV-----IAKNGDL 844
Db 872 NANTAIDQLPNLNQPKTALKQDVSHAELVTGVNAIKQADALNNAMGTLKQIQIANSQV 931
QY 845 FKSNGITQFEMAPQYVSSEDCGTFDLSVILNG-----YAFSDRYDLAKSKKNYKSGQDLAN 900
Db 932 POSVDFTQADQDKQQAAYNNAQAQQ-IANGIPTVPLTPTVTQAVTTNQ-----AKDALN 987
QY 901 AIKGLQSAQIKVLSDLVPNQLPGKEVVVTAFRVQYQQAQSGATINTKTPY----- 952
Db 988 GDEKLAQAKQEAALANLDTLRDLNQPORDALR-----NQINOQAALATVEQTKQNAQNVNTA 1043
QY 953 -----VAN-----TRSYGDYQEQYGGKFLDQLKYLPRLFSTTKQISTGKPIDPSV-- 997
Db 1044 MSNLKQGIANKDQTVKASENVHDADA-----DKQTAYTNVSAQEGIIQNTNPTLNP 1095
QY 998 -KITNWSAKYFNGSNILGRGAKYVLSEGN----- 1025
Db 1096 DEITRALTOVTDKAKNGLNGEAKLATEKQNAKDAVSGMTHLNDQAQKALQKQIDQSPEIAT 1155
QY 1026 -----KVLNLADGKFLPTVLN-----NTYGOQVQSANGF 1055
Db 1156 VNQVKQTATSLQDMQLSQALNDKQATLDG-----NYLNADPPDKQNAKYQAVAKAEAL 1210
QY 1056 ISKNGSIHYLDKNGQEVKNRFEISGSWYTFYFSDGKQAT-----GKTKIGINDTYL 1105
Db 1211 LNKQSGTNEVQAVESITN--EVNAAKQALNGNDNLNAAKQQAQKQALANLTHL 1261

RESULT 15
US-11-013-759-9
; Sequence 9, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loomsore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:Jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2053
; TYPE: PRP
; ORGANISM: Moraxella catarrhalis
; US-11-013-759-9

Query Match      3.2%; Score 231; DB 7; Length 2053;
Best Local Similarity 19.9%; Pred. No. 1e-05;
Matches 297; Conservative 187; Mismatches 573; Indels 438; Gaps 72;

QY 27 SFLAGQVVVAADTNNDGTSVQVNNKVPSPDKFDAQONGQALQAQAFKAAQADQATATSQ 86
Db 515 STLNSGSLTVNNTTGNKQIQVGANGI-----KFATVANN-----VANTSATVGTAR 560
QY 87 VS-----PATDGRVDNQTTPAANQPAANVANQDVANPATDAG--ALANQSAADT-----S 134
Db 561 ITEEKIGFAGTNDGVDEQ-APYLDKERLKVGRVEI---TTDSGINAGNHKITGLTNGIAN 616
QY 135 TDGKAVPQTSQPGHLETVGKTYVVDVANGQRLKNYSMVIDGKTYVFDGTEGAEATDLPK 194
```

Db 617 TDAVTKQLKDAKPTLNAGDGLS--INSNNGDLVDSSGNIITPTTNYISVKT-----TKLNS 670
Qy 195 TQOANODNVPDSYQANNOAYSNEASFETVDNYL-----TADSWYRPRKILKNQSGHWAAS 249
Db 671 NGTSGN---NKFSVSNADHNSLVTAKDLADYLNKVNETAADALSALPSFKVQNGDSSNNAI 726
Qy 250 SEGDLRPLMTWPDAAATKAAAYANFWAKEGLISGSYRQNSANL--DAATQNTQSAIEKK- 306
Db 727 TVG-----KDTNGKTFNTLKLKG-----ENGVNITTRATGTVTFGIDQSN 767
Qy 307 -----IASBGTNWLNRDKMSQFVKSONQMSIASENETVYPNQDHMOGGALLFNS-- 356
Db 768 GLTTPKLTVCSTNGNRLVIEQVPSADGNSKTNIIKGLSPTLPSIASPSGRNIALGNTIE 827
Qy 357 -KDEHANSDBWLLNRNPTFO-TGKQKYFTTNAGYELLANDVNSNPNVQAEQLNHLH 414
Db 828 EKDKSNAASIDVDLNAAGFLNKNKDKDFVSTYDVTVDFTDGNATTATVYDEANQTSKVA 887
Qy 415 YLNNWGDIVM---GDKDANPDGVR---VDAVDNVNADLLQIORDYYKAK----- 457
Db 888 YDVNVDEKTIELTGDNGKKQLGKVTIKLTETSTNGNATTFSTDDHALVKASDIAGNLMT 947
Qy 458 -----YGT-----DONKNAIDHLSIL---EAWSGNDNDY-VKDONNFS 492
Db 948 LAEEIHTTKGTANTALQTTVKKVBENDK-ADDTNATVKGDKGTSGKVNLTCLKKNGGLD 1006
Qy 493 LSIIND-----QRSGMLKAFGYASAYRGNLSNLATAGL--KNRSANPD-----SDPVP 538
Db 1007 IKTRDKGTVTGINTQSG-LKA-----GDSSTLANNGLSIKNTASNEQIQVGADGVK 1057
Qy 539 NVVFRADHSEVQTRI---AKIIREKLGKTNADGLTNLTDLNLKAFDIYNQDMNATDKV 595
Db 1058 ---FAMVNNGVGAGIDGTRITRDEIGFTGTNGSLDKSKPHLSK-----DCINAGGK- 1107
Qy 596 YYPNNLPMAYANMLQKDTVRVYVGYDMYTDNGQYMATKTPFYNALETLLKGRIKYVAGG 655
Db 1108 -----KITNIQSG-----IAKNSHDAVTGG 1128
Qy 656 QAVSYNQDWSGILTSVRYKGA-----NSASDAGNTETRNSGMALLINRN-PRAY 707
Db 1129 KIYDLKTELENKISSAKTAQNSLHEFSVADEQGNFTVSNPVSSYDTSKTSDDVITFAGE 1188
Qy 708 RMLT--LNGAAH-----KSOAYRPLLLSTKDGATYLANDSDVRSROYKYTDSQNLSP 759
Db 1189 NGITTKVNRGVVRVIGIDQKGLTTPKLTGVNNGKGIVINSQ-----NGQNTITG 1238
Qy 760 SASELQSVANAQVSGMIQVWPVPGAADNQDVRTSPSTQATKDCNIIYHQSD-----ALDSQV 815
Db 1239 LSNLTANVTN-----DKGSVRT-----TEQNIILKDEKTRAASIVDV 1276
Qy 816 IYEGFSNFOAQPSPQDYT-----NAVIK--NGDLPKSWG------TQFEM 855
Db 1277 LSGAF-NLQNGEAVDFVSTYDTPNFANGNTTTAKVYDDTSTKTSKVYDVNVVDDTTIEV 1335
Qy 856 AFQYVS-----SEDTGLDSVLINGYAFSDRY-----DLAMSKNKYKSGQDLANAIGL 905
Db 1336 KDKKLGKVTTLTSTGTGANKFALSNOATGDALVKAASDVAHLNLTLSGDIQAKGASQAN 1395
Qy 906 QSAGIKVLSDLPNQLNPLPGKEVVTATRVNOYQAKSGATINKTPYVANTRSYGDYQEQ 965
Db 1396 NSAG-----YVDAGNKVIYDSTDNKYIQAANDGTVDKTEVAKDKLVAAQATP 1444
Qy 966 YGCKFLDDILQKLYPRLFSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSEGN 1025
Db 1445 DG-----TLAQM-----NVKSVINKEQVNDANK-----KQGINEDNAFVKGLEKAASD-N 1488
Qy 1026 KYLNADGKLFPLTV--LNTYGOPOVSA--NGFISKNGGIHYLDKNGQEVKVRKFE--- 1078
Db 1489 KTKNAA-----VTVGDLNAVAQTPLTFFAGDTGTAKLGETLTIKGGQDTNKLTDNNI 1542
Qy 1079 --ISGSWYVDSDG---KMATGKTKIGNDTYLPMFMPNGKQKKEGVWYDGKKAIVYDDNGRT 1133
Db 1543 GUVVAG-----TDGFTVKLAKOLTNLSVN-----AGGTKIDEXISFVDANGQA 1586

Qy 1134 WTNKGFVEFRVDGQDKWRYFNGDGTIAIGLVSLDNRLTLYFDAYGYQVKGTQVTINGKSYT 1193
Db 1587 KAN-----TPVLSANGLDLGGKVISNVGKG-T 1612
Qy 1194 FDADQGLVQTDNANPAPQGOAGWKLLG---DNQWGYR-----K 1229
Db 1613 KDTDAANVOQLNEVR-----NLLGLGNDNADGNQVNIADIKKDPNSGSSSNRTVIK 1663
Qy 1230 DQQLLTGBQTTIDGQKVFFQDNGVQV---KGGTAT--DASGVLRFYDRDQGHQVKGWYSTS 1285
Db 1664 AGTVLGGKGNNDTEKL--ATGGEVQGVVDKGNANGDLSNV-----WVKTQ 1706
Qy 1286 DDNWWYVNESGQVLTGLQTTIDGQTYFFDDKGIQAKGKAVWDENGNLRYFDADSGN 1340
Db 1707 KDG-----SKKALLATYNAAGQTYVTNNPABAIDRI--NEQG-IRFPHVNDGN 1752

Search completed: February 11, 2006, 20:58:50
Job time : 18.7677 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 19:15:50 ; Search time 28.1742 Seconds
(without alignments)
4661.567 Million cell updates/sec
Title: US-10-797-821-39
Perfect score: 7230
Sequence: 1 MEKNLYKLHKVKKQWVAIG.....WKNVDGNWYFNRNGLATRW 1365
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR 80:*
1: piri:*
2: piri2:*
3: piri3:*
4: piri4:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7230	100.0	1365	2 A41483	glucosyltransferas
2	3174	43.9	1431	2 A45866	dextranucrase (EC
3	3079.5	42.6	1449	2 T30857	glucosyltransferas
4	3076.5	42.6	1449	2 T30552	glucosyltransferas
5	3026.5	41.9	1577	2 T30858	glucosyltransferas
6	3023	41.8	1592	2 A38175	glucosyltransferas
7	2945.5	40.7	1475	2 B33135	gtfB protein precu
8	2893.5	40.0	1375	2 J20345	dextranucrase (EC
9	2863.5	39.6	1599	2 S22737	glucosyltransferas
10	2804	38.8	1518	2 A44811	glucosyltransferas
11	2702	37.4	1508	2 T31098	probable dextranu
12	2593.5	35.9	1290	2 JCS473	dextranucrase (EC
13	543	7.5	2817	2 B97033	uncharacterized pr
14	367.5	5.1	2710	2 A37052	toxin A - Clostrid
15	358.5	5.0	2364	2 I40884	cytotoxin L - Clo
16	334.5	4.6	2367	2 S70172	toxin B - Clostrid
17	327.5	4.5	2366	2 S10317	toxin B - Clostrid
18	323	4.5	2178	2 S55805	alpha-toxin - Clo
19	316.5	4.4	563	2 A37184	glucan-binding pro
20	306.5	4.2	2334	2 S32920	cell wall-associat
21	285	3.9	648	2 S12920	enterotoxin A - Cl
22	273	3.8	1441	2 B86807	hypothetical prote
23	270.5	3.7	690	2 F98114	choline-binding pr
24	267	3.7	2167	2 AF1489	cell wall-associat
25	260	3.6	1999	2 AB2018	hypothetical prote
26	258.5	3.6	1385	2 D89824	hypothetical prote
27	256	3.5	1463	2 T30290	AAS surface protei
28	247.5	3.4	2703	2 H81193	hemagglutinin/hemo
29	247	3.4	583	2 S57721	cspB protein - Clo

ALIGNMENTS

RESULT 1

A41483
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Mar-2004
C:Accession: A41483
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
Infect. Immun. 58, 2452-2458, 1990
A:Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltran
A:Reference number: A41483; MUID:90316665; PMID:2142479
A:Accession: A41483
A:Molecule type: DNA
A:Residues: 1-1365 <GIL>
A:Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:g153652; PIDN:AAA26898.1; PID
C:Genetics:
A:Gene: gtfS
C:Keywords: glycosyltransferase; hexosyltransferase
Query Match 100.0%; Score 7230; DB 2; Length 1365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEKNLYKLHKVKKQWVAIGVTTVTLSTFLAGQVVAADTNNNDGTSVQVKNVPSDPKFD 60
DB 1 MEKNLYKLHKVKKQWVAIGVTTVTLSTFLAGQVVAADTNNNDGTSVQVKNVPSDPKFD 60
QY 61 AQAQNGQLAQAAMFKAAQADQATTSQVSPATDGRVDNQVTPAANQPAANVANQDVANPAT 120
DB 61 AQAQNGQLAQAAMFKAAQADQATTSQVSPATDGRVDNQVTPAANQPAANVANQDVANPAT 120
QY 121 DAGALNROSADTSTDGKAVPQTSPOGHLETVDGKTYVVDANGQRLKNYSWIDGKTYT 180
DB 121 DAGALNROSADTSTDGKAVPQTSPOGHLETVDGKTYVVDANGQRLKNYSWIDGKTYT 180
QY 181 FDGQGEAQTDLPKTGOANQDNVPSYQANNQAYNEASSFETVDNYLTADSWYPRKIL 240
DB 181 FDGQGEAQTDLPKTGOANQDNVPSYQANNQAYNEASSFETVDNYLTADSWYPRKIL 240
QY 241 KNGSQWQASSEGDLRPIPLMTWPPDAATKAAVANFWAKEGLISGSYRQNSANLDAATQNIQ 300
DB 241 KNGSQWQASSEGDLRPIPLMTWPPDAATKAAVANFWAKEGLISGSYRQNSANLDAATQNIQ 300
QY 301 SAIKKIASEGNTWLRDKMSQFVKSNQWSTASENETVYPNQDHMQGALLFNSKQTE 360
DB 301 SAIKKIASEGNTWLRDKMSQFVKSNQWSTASENETVYPNQDHMQGALLFNSKQTE 360
QY 361 HANSDWRLNRPFTQTKQKYFTTNAGYELLANDVNSNPVYQAEQNLHLYMNGW 420
DB 361 HANSDWRLNRPFTQTKQKYFTTNAGYELLANDVNSNPVYQAEQNLHLYMNGW 420
QY 421 DIVMGDKDANFDGVRVDVNVNADLLQIQRDYYKAKYGTQDQNEKNAIDHLSILEAWSGN 480
DB 421 DIVMGDKDANFDGVRVDVNVNADLLQIQRDYYKAKYGTQDQNEKNAIDHLSILEAWSGN 480

Db 421 DIVMGDKANFDGVRVDAVDNVNADLLQIQRDYKAKYCTDQNEKNAIDHLSILEAWSN 480
QY 481 DNDYVKQDNFSLSDNDORSGLMKAFGYASAYRGNLSNLTAGLKNRSANPDSPVPNY 540
Db 481 DNDYVKQDNFSLSDNDORSGLMKAFGYASAYRGNLSNLTAGLKNRSANPDSPVPNY 540
QY 541 VFIKRAHDSVOTRIAKIIEKLGKTNADGLTNLTDLDLNAKADPIYNQDMNATDKVYPPN 600
Db 541 VFIKRAHDSVOTRIAKIIEKLGKTNADGLTNLTDLDLNAKADPIYNQDMNATDKVYPPN 600
QY 601 LPMAYAWMLQNKDVTTRVYGYDMYTDNGQYMATKTPFYNAIETLLKGRIKYYVAGQAVSY 660
Db 601 LPMAYAWMLQNKDVTTRVYGYDMYTDNGQYMATKTPFYNAIETLLKGRIKYYVAGQAVSY 660
QY 661 KQWSSGILTSVRYKGGANSASDAGNTETRNSGMALLINRPNFRAYRNLTLNMGAAHKS 720
Db 661 KQWSSGILTSVRYKGGANSASDAGNTETRNSGMALLINRPNFRAYRNLTLNMGAAHKS 720
QY 721 QAYRPLLSTKGIATYLNDSVDVSRQYKTYDSQGNLSPSASELOSANAOVSGMIOVWV 780
Db 721 QAYRPLLSTKGIATYLNDSVDVSRQYKTYDSQGNLSPSASELOSANAOVSGMIOVWV 780
QY 781 PVGAADNQDVRTSPSTQATKQNIYHQSALDSQVIYEGFSNFQAFQSPDOYTNVIAK 840
Db 781 PVGAADNQDVRTSPSTQATKQNIYHQSALDSQVIYEGFSNFQAFQSPDOYTNVIAK 840
QY 841 NGDLFKSWGITQFEMAPQVVSSEDTGFLDSVILNGYAFSDRYDLAMSKNKGYSKODLAN 900
Db 841 NGDLFKSWGITQFEMAPQVVSSEDTGFLDSVILNGYAFSDRYDLAMSKNKGYSKODLAN 900
QY 901 AIKGLQSGIKVLSLVPNQLNLPKGVVTVATRVNOYGOAKSGATINKTPVANTRSYG 960
Db 901 AIKGLQSGIKVLSLVPNQLNLPKGVVTVATRVNOYGOAKSGATINKTPVANTRSYG 960
QY 961 DYQEQYGGKFLDLQKLYPRLFTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKTV 1020
Db 961 DYQEQYGGKFLDLQKLYPRLFTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKTV 1020
QY 1021 LSEGKNYLMADGKFLPVLVANNYGPQVVSANGFISKNGGIHYLDKNGQEVKNRPFKEIS 1080
Db 1021 LSEGKNYLMADGKFLPVLVANNYGPQVVSANGFISKNGGIHYLDKNGQEVKNRPFKEIS 1080
QY 1081 GSWYFDSGKWATGKTIGNDTYLPMNGKOLKEGVYDGGKAYVYDDNGRTWTKGFV 1140
Db 1081 GSWYFDSGKWATGKTIGNDTYLPMNGKOLKEGVYDGGKAYVYDDNGRTWTKGFV 1140
QY 1141 EFRVDGQKRWYFNGDGTIAIGLVSDNRTLYFDAYGYQVKGQVTVINGKSYTFDADQGD 1200
Db 1141 EFRVDGQKRWYFNGDGTIAIGLVSDNRTLYFDAYGYQVKGQVTVINGKSYTFDADQGD 1200
QY 1201 LVQTDNANPAQOQAGKLLGNQNGYRKDGQLLTGEBQITIDGQVFFQDNGVQVGGTAT 1260
Db 1201 LVQTDNANPAQOQAGKLLGNQNGYRKDGQLLTGEBQITIDGQVFFQDNGVQVGGTAT 1260
QY 1261 DASGVLRFYDRDQGHQVKGWYSTDNNWYVNESQVLTGLTIDGQTVYFDDKGIQAK 1320
Db 1261 DASGVLRFYDRDQGHQVKGWYSTDNNWYVNESQVLTGLTIDGQTVYFDDKGIQAK 1320
QY 1321 GRAVDENGLRYFDADSGNMLRDRKXNDGNWYFNRRNGLATRW 1365
Db 1321 GRAVDENGLRYFDADSGNMLRDRKXNDGNWYFNRRNGLATRW 1365

RESULT 2
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866
J.Honda, O.; Kato, C.; Kuramitsu, H.K.
A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A:Reference number: A45866, MUID:91100958, PMID:2148600

A:Accession: A45866
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 43.9%; Score 3174; DB 2; Length 1431;
Best Local Similarity 47.3%; Pred. No. 2.4e-153;
Matches 677; Conservative 213; Mismatches 436; Indels 104; Gaps 38;

QY 1 MEKNRYKLVKKKQWVAIGVTT--VTLSFLAGGOVVAADTNNNDGTSV-----QVNKAV 53
Db 1 METKRYKWHKVKHVVTVAVASGLITLTGTTGLGSSVSAETEQQTSDKVVTOKSDDKAA 60

QY 54 PSDPKFDA---QAQNGQLAQAMFKAANOADOTATSOVSPATDGRVDNQVTPAANQPAAN 109
Db 61 SESSQTDAPKTKAQTEQ-TQAQ-SQANVAD-TSTIS-----ITKETPSQNIITQANSDDKT 113

QY 110 VANQDVANPATDAGALNRQSAADTSTDGKAVP-----QTSQDPG-----H 149
Db 114 VTNKSEEAQTSSEERTKQSEEAQTASSQALTQAQAEILTQKQRTAAQENKNPVDLAAIPN 173

QY 150 LETVDGKTYVDANGORLKNYSWIDGKTYTFDGTGEAQTDLP----KTQANODNVPD 205
Db 174 VKQIDGKTYTGSQDQPKKFNALTVNNKVLVPDKMTG-ALTDTSYQFQKQGLTKLNN--- 229

QY 206 SYQANNOAYSNEASFETVDNYLTADSWYRPKILKNGOSWOASSEGDLRPLIMTWPPDA 265
Db 230 DYTPHQVQINVENTSLETIDNYVTADSWYRPKDLKNGKWTASSEDLRPLIMSWWPK 289

QY 266 ATKAAAYANFWAKEGLISG-SYRQNSA--NLDAATQNIQSAIEKKIASEGNTWLNRKMSQ 322
Db 290 QTOIAYLVNMQOGLTGENTYADTSQESLNLAQTQVVKIETKISQTOQTQWLARDIINS 349

QY 323 FVKSONQHSIASENETVPNQDHMQGALLPSNSKDTSEHANSKDWELLARNPTFTGKQKY 382
Db 350 FYKTOPNWSQTESDTSAGKDLQGLGALLYSNSDKTAYANSDYRLNRTPTSTQYKPKY 409

QY 383 FTTNYA-GVELLLANDVNSNPVQAEQLNHLHYLMNWGDIVMGDKDANFDGVRVDAVDN 441
Db 410 FEDNSSGGYDFLLANDINDSNPVPVQAEQLNHLHYLMNYGSIIVANDPEANFDGVRVDAVDN 469

QY 442 VNADLLQIORDYKAKYGTQNEKNVAIDHLSILEAWSNDNDYVKQNNFSLISINDQRS 501
Db 470 VNADLLQIASDYLKARYGVDKSEKNAINHLSILEAWSNDNDPOYNKDTKGAQLPIDNKLRL 529

QY 502 GMLKAF-----GYASAYRGNLSNLTATAGLKNEA-NPDSDPVNTYVFIKRAHSEVOT 552
Db 530 SILLYALTPLKADSNKNEIRSGLEPVTNLSNNNSAEGKNSERWANYIFIKRAHSEVOT 589

QY 553 RIAKIIRKLGKTNADGLTNLTDLDLNAKADPIYNQDMNATDKVYPPNPLPMAYAWMLQNK 612
Db 590 VTAKIICAQI-NPKTDTGLT-FTLDELKQAFKIYNEDMRQAKKYYQTSNIPTAYALMLSNK 647

QY 613 DTVTRVYGYDMYTDNGQYMATKTPFYNAIETLLKGRIKYVAGQ--AVSYKO-----DW 664
Db 648 DSITRLYYGYDMYSDGQYMATKSPYYDAIDTLTKARIKYAAGGQDMKITYVEGDKSHMDW 707

QY 665 S-SGLTTSVRVCKGANSASDAGNTRNSGMALLINRPNFRAYRN--LTLNMGAAHKSQ 721
Db 708 DYTGLVTSVRVGTGANEATDQGEATKQGMVAVITSSNNPSLKNQNDKIVVNMGAHKNQ 767

QY 722 AYRPLLSTKGIATYLNDSVDVSRQYKTYDSQGNLSPSASELOSANAOVSGMIOVWV 781


```
Db 768 EYRPLLLTTKDGLTSTSDAAKS-LYRKTNDKGELVFDASDIQGYLNPQVSGYLAVWP 826
Qy 782 VGAADNQDVRTSPSTQATKDNIGYHOSDALDSQVIYEGFSNFQAFQSPDQYTNVIAKN 841
Db 827 VGASDQDVRVAASNAKATQGVYESSALDSQLIYEGFSNFQDFTVKDSYTNKKIAQN 886
Qy 842 GDLFKSWGITQFEMAFQYVSSEDTGLDLSVLNGYAFSDRYDLAMSKNNKYKSKDLANA 901
Db 887 VOLFKSWGVTSEFMAFYVSSEDSGLDSIIQNGYAFEDRYDLAMSKNNKYKSKQDMINA 946
Qy 902 IKGLQSGIKVLSDLPNOLYNLPKGEVVTATRVNOYGOAKSGATINKTPYVANTRSYG- 960
Db 947 VKALHSGSQIADWPDQIYNLPKGEVVTATRVNDYGEYKDSSEIKNTLYAANTKNGK 1006
Qy 961 DYQEQYGGKPLDQLKLYPRLFTSKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYV 1020
Db 1007 DYQAKYGGAPLSLAAYKPSIFNRTQISNGKIDPSEKITAWKAKYFNGSNILGRGVYV 1066
Qy 1021 LSE--GNKYNLADGKLFPLTVLNNYTGQPOVSANGFISKGGIHYLDKNGQEVKRF-K 1077
Db 1067 LKDNASDKYFELKGNQTYLPKQMTN-----KEASTGFVNDGNGMTFYSTSGYQAKNSFVQ 1121
Qy 1078 EISGSWYFSDSKMATGKTIKNDTYLFPNGKQLKEGVW--YDGGKAYYYDDNGRTWT 1135
Db 1122 DAKGNWYFDNNGHMYGLQNGEVOYFLSNGVQLRESFLENADGSK-NYFGLGNRYS 1180
Qy 1136 NKGFEVFRYDGDQKRYFNGDGTIAIGVSLDNRTLYFDAYGYQVKGQTVT-INGKSYTF 1194
Db 1181 N-GYYSF--DNDSKRYFDASGVMAVGLKTNGTQYFDQDGYGVKGAMITGSDGKKRYF 1237
Qy 1195 DADQGLVOTDNANPAQOAGKWLGDNQWGY-RKDGQLLTGEBQITDQKVFQDNGVQ 1253
Db 1238 DDCSGNMAVNFANDK-----NGDWYLYNSDGIALGVQVQINGKTYFYFGQDGKQ 1286
Qy 1254 VKGSGTADASGLVRFYRDQGHQVKGWYSTDDNMYVNESQVLTGLQITDQVTVYED 1313
Db 1287 IKGKIITD-NGKLYFLANGELARNIFATDSNNWYFSGDVAVTGSQTTAGKLYFA 1345
Qy 1314 DKGIOAKGKAVMDENGLRYFDADSGNMLRDRWK-NVDGNWYFNRNGLA 1362
Db 1346 SDGKQVKGFSV-TYNGKVHYHADSGELQVNRFEADKQGNWYLYDSNGEA 1394

RESULT 3
T30857
Glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30857
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: Z20909, MUID:95122197, PMID:7822030
A:Accession: T30857
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166E; EMBL:L35495; NID:g6623378; PID
C:Genetics:
A:Gene: gtfI

Query Match 42.6%; Score 3079.5; DB 2; Length 1449;
Best Local Similarity 45.0%; Pred. No. 1.6e-148;
Matches 667; Conservative 206; Mismatches 451; Indels 157; Gaps 35;

Qy 1 MEKRLRYKLHKVKQKQVIAVGVTTVLTSFLAGGQVVAADNNNDGTSVQVKNKWPSPDKPD 60
Db 1 MDKRVHYKMKVKQKQVTVIAVGLSL-----GAVSAVSLGTNDG-----VVQADEHTD 48
Qy 61 AQAQNGQLAQAQKFAANO--ADQATATSOVSPATDGRVDNQVTPAA-----NOPAAN 109
Db 49 ATVAIPDITVDRTGTVNSDPTTAQDPTTAAATNDVATD-QATPTATFDLTDTTNTVAAN 107
```

```
Qy 110 VANQDVANPATDAGALNRQSA-----DTSTDGKAV----- 140
Db 108 -AVDTVATVGTDRAAATTNDTATNDTAVDTNNNTTDTTVDTRAATTERRATGARRGP 166
Qy 141 -----POT-----SDQPHLETVDGKTYVVDANGQBLKN 169
Db 167 TGRRRATPVNGTNNANNTVTVVNNDLPATNNVVTDPGSHIKTINGKQYVVEDGTIRKN 226
Qy 170 YSNVIDGKTYVFDQGTGEAQTD----LPKTGOANQDNVPDSYQAN-----NOAYSNEA 218
Db 227 YVLERIGGSOYFNAETIGELSKQYRFDKNGGTGSS--ADSTNTNTVNGDKNAFYGTTD 284
Qy 219 SPSFETVNYLTADSWTRPRKILKNGOSWQASSEGDLRPIILMTWMPDAATAKAAAYANFAWE 278
Db 285 XDIELVDGYFTANTWYRPEIKLKGKWEASTASTENDKRPLLTVMWPSKAQASLYLTMKEQ 344
Qy 279 GL-ISGSYRQNSA--NLDAATQNIQSAIEKKIASEGNTNMLRDKMSQFVKSONQWIAS 335
Db 345 GLGTNTQYTSFSQTMQOAALEVQKRIERAREGNTDMLRTTIKRVKTOPGMNSTSE 404
Qy 336 NETVTPNQDHMOGGALLFNSKDTTEHANSDFWLLNRNPTFQTKQ--KYP-TTNYAGYEL 392
Db 405 N--LDNNDHLOGGALLYNNDSRTSHANSDYRLARTPTSQGKNPKYTKDTSNGGFEP 461
Qy 393 LLANDVNSNPVVAQBLNHLHYLMNWGDIVMGDKDANFDGVRVADVADNVNADLLQIQRD 452
Db 462 LLANDIDNSPVAQBLNHLHYIMNIGTITGSEDEDFGVRVADVADNVNADLLQIASD 521
Qy 453 YKAKYGTQDNEKNAIDHLSIILEANGSNDNDYKQDNFSLSDNDORSGLKAFGYASA 512
Db 522 YPKAKYGADQSDQAKHLSIILEANSHNDAYYNEDTKGAQLPMDPMHLALVYSLLRPIG 581
Qy 513 YRGNLSNLATAGLKNRS-ANPDSDPVPNVYFIRAHDSVQVTRIAKIREKLGKTNADGLT 571
Db 582 NRSQVEPLISNLSNDRSESGKSKRMANYAFVRAHDSVQSTIGQIIKNEINPQSTG--N 639
Qy 572 NUTLDDLKKAFDIYNQDMNATDKVYYPNNLPMAWMLQNKDQTVTRVYVYGDYTDGQYM 631
Db 640 TFTLEMKKAFBIYNKDMRSANKQYQYNIPSAYALMLTKDQTVPRVYVYGDYTDGQYM 699
Qy 632 AKTTPFYNAIETLLKRIKYVAGGQ--AVSY-----KQDW-SSGILTSRYRKGANSASD 683
Db 700 AQKSPYDAIETLLKRIRYAAGGQDMKVNITGYGNTNGMDAAGVLTSTVRYGTGANSASD 759
Qy 684 AGNTESTRNSGALLNNRPNFRAYRNLTLNMGAAKHSQAYRPLLLSTKDGIAITYLNDSDV 743
Db 760 TGTATRNQGMVIVSNQPALRLTSLNLTINMGAAHNRQAYRPELLTTNDGVATYLNDSA 819
Qy 744 DSRQYKTYDSQGNLSFSASELOSANAVQSGMIQVWVPVGAADNQDVRTSPSTQATKGN 803
Db 820 NG-IVKYTDGNGNLTFSEANEIRGNPQDGYLAVWVPVGAASENQDVRVAPSKENSSGL 878
Qy 804 IYHQSADLDSQVIYEGFSNFQAFQSPDQYTNVATKNGDLPKSWGITQFEMAFQVVSSE 863
Db 879 VYESNAALDSQVIYEGFSNFQDVFQNPQSYTNKKAENANLFSKSWGITSFEFAPQVSSD 938
Qy 864 DGTFLDLSVLNGYAFSDRYDLAMSKNNKYKSKQDLANAIKGLQSGIKVLSDLPNOLYN 923
Db 939 DGSFLDSVLQNGYAFDTRDYIGMSKDNKIGSLADKAALKSLHAGVISAIADWVPDQIYN 998
Qy 924 LFGKEVVTATRVNOYGOAKSGATINKTPYVANTRSYG-DYQEQYGGKFLDLDLQKLYPRL 982
Db 999 LFGDEVVTATRVNNYGETKQGAIIHSLYAAKTRTFGNDYQYKYGAGFLDELKRLYPOIF 1058
Qy 983 STKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSSEG-NKYNLNLADGKLFPLTVL 1041
Db 1059 DRVQISTGKRMPTDEKITQWSAKYMNVTNILDGSEYVLKNGLNGYGTNGGKVSUPKV 1118
Qy 1042 NNTYQGPQVPSANG-----FISKNGGIIHYLDKNGQEVKQRF-XEISGSWYFSDGKMWATG 1095
Db 1119 GSNQSTNGDNQNGDGGSGKPEKRLFSVRYRNNNGYAKNAFINKDNGNVYIFDNGSMVAG 1178
Qy 1096 KTKIGNDYLFPNGKQLKEGVWYDQK-KAYYYDDNGRTWTN-----KGFEV 1141
```

Db	1179	EXTIDGKQYFFLANGVQLDGYRQNRGQVFFYDQNGVLNANGKQDPKPDNNNNSGRNQ	1238
Qy	1142	FRVDGQDKWRYFNGDGTIAIGLVSLDNRITLYFDAYGVQKQTVTINGKSYTFDADQDGL	1201
Db	1239	FVQIGNVWAYDYGNGKRVTHQNINGQBELFFDNNNGVQVKGRTVNGGAIRYDANSSEM	1298
Qy	1202	VQTDNANPAQOAGKWLGDNQWGY-RKDGQLLTGEGTIDGQKVPFDQNGVQVKGTTAT	1260
Db	1299	ARNRFAETEP-----GVWAYFNNDGTAVKGSQNINGQDLYFDQNGRQVKGALA-	1346
Qy	1261	DASGVLRFYDRDQGHQVKGWSTSDNNWVYVNESGOVLTLQTLTIDGQTVYFDDKGIQAK	1320
Db	1347	NVDGNLRYDVNSG-ELYRNRPHEIDGSWYFDGNGNAVKGWNINGQNLFFDNNKGQIK	1405
Qy	1321	GKAVWDEGNLRYFDADSGNMLDRWKNDV-GNWFYFNRNG	1360
Db	1406	CHLV-RVNGVVRVYFDPNSGEMAVNRWVEVSPGWVYFDGEG	1445
RESULT 4			
T30552			
glucosyltransferase N - Streptococcus salivarius (fragment)			
C:Species: Streptococcus salivarius			
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T30552			
R:Jaife, R.I.			
submitted to the EMBL Data Library, February 1998			
A:Description: Streptococcus salivarius V1477 gtfN.			
A:Reference number: Z20854			
A:Accession: T30552			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1449 <JAF>			
A:Cross-references: UNIPROT:O68542; UNIPARC:UPI0000B10FD; EMBL:AF049609; NID:G2935545;			
C:Genetics:			
A:Gene: gtfN			
Query Match 42.6%; Score 3076.5; DB 2; Length 1449;			
Best Local Similarity 45.0%; Pred. No. 2.2e-148;			
Matches 667; Conservative 205; Mismatches 452; Indels 157; Gaps 35;			
Qy	1	MEKNRLYKLVKQKQWVAIGVTVTLTSLFAGQVVAADTNNDGTSVQVKNVPSDPKPD	60
Db	1	MDKKVHKVHKVKKQWTVAVTGLSL-----GAVSAVSLGTNDG-----VQADEHTD	48
Qy	61	AOAQNGQLAQAMFKAANQ--ADQTATSVQSPATDGRVDNQVTPAA-----NQPAAN	109
Db	49	ATVAIPDIIVDTGTVSNDTTAAQDPTTAAATNDVATD-QAFTATFDLTDTTNTVAAN	107
Qy	110	VANQDVANPATDAGALNRQSA-----DTSDDKAV-----	140
Db	108	AVDTVATVGTDRAAATNTDTATNDTAVDTNNNTTDTTNTNRAATERRATGARRGP	166
Qy	141	-----POT-----SDPGHLETVGKTVYVDANGORLKN	169
Db	167	TGRRATPVNGNTNANNVTVVVNDLPATNVVTDGSPHIKTNGKQYVVEDDGTIRKN	226
Qy	170	YSKWIDGKTYFDGQTEGAQTD-----LPKTQGANQDNVPSYOAN-----NQAYSNEA	218
Db	227	VYLERIGGSQYFNAETGELSNOKEYRFDKNGGTGS--ADSTNTVTVNGDKNAFYGTTD	284
Qy	219	SSFETVDNYLTADSWYRPKRILKNGSQWASSEGDLRPLMTWPDPAATKAAAYANFWAKE	278
Db	285	KDIELVDGYFTANTWYRPKEILLKDGKEWTASTENDKRPLLTVMWPSKATQASYLYNMYKEQ	344
Qy	279	GL-ISGSYRQNSA--NLDAATQNIQSAIEKKTASGNTWMLRDKMSQFVKSQNSIASSE	335
Db	345	GLGTNTQYTSFSQTMQDAALEVQKRIEGRAREGNTDMLRTIKNFVKTPQFGWNSTSE	404
Qy	336	NETVVPNODHMOGALLFNSKDTTEHANSDWRLLRNPTFTQCKQ--KYF-TTNYAGYEL	392
Db	405	N---LDNNDHLQGGALLYNNDSRTSHANSYRLLRNPTTSQTGKINPKYTKDTSNGGPF	461
glucosyltransferase - Streptococcus salivarius			
RESULT 5			
T30858			
glucosyltransferase - Streptococcus salivarius			
Qy	393	LLANDVNSNPVQAEQNLHLHYLNMNGDIIVMGDKDANFDGVVRVDAVDNVNADLLQIQIR	452
Db	462	LLANDIDNSNPAVQAEQNLWLYHINIGTITGGSEDENFDGVRVDAVDNVNADLLQIQIR	521
Qy	453	YKAKYQTDQNEKNAIDHLSILEAWSGNDNDYVDQNNFSLSIDNDORSGLMKAFGYASA	512
Db	522	YFKAKYGADQDQQAIAKHLHLSILEASHNDAYNEDTKGAQLPMDDPMHLLAVYSLLRPIG	581
Qy	513	YRGNLSNLTATAGLKNRS--ANPDSDPVPVNYFIRAHDSVQVTRIAKIIREKLGKTNADGLT	571
Db	582	NRSQVEPLISNLSLDRSSGKNSKEMANYAFVRAHDSVQSIIGQIIKNEINPQSTG--N	639
Qy	572	NLTLDLKAEDIYNQDNATDKVYYPNNLPMAVAMLONKDTTVRVVYGYDMYTONGQYM	631
Db	640	TFTLDKMKAFEIYNQDNATDKVYYPNNLPMAVAMLONKDTTVRVVYGYDMYTONGQYM	699
Qy	632	ATKTPFYNAIETLLKGRIKYVAGQ--AVSY-----KODW--SSGILTSRVYRGKANSASD	683
Db	700	AQKSPYDAIETLLKGRIRYAAGQDMKVNYIGYNTNGWDAAGVLTSVRYGTGANSAD	759
Qy	684	AGNETRNSGALLINRPNFRAYNLTNLNGAAHKSQAYRPLLLSTKQGIATYLNDSV	743
Db	760	TCTAETRNQGMVIVSNQPALRLTSLNLTINMGAAHQAHRNQAYRPLLLTTNDGVATYLNDSA	819
Qy	744	DSROVKYTDSONLSFSASELSQVANAQVSGMIOVWVPVGAADNODVRTSPSTQATKQGN	803
Db	820	NG-IVKYTDGNGNLTFSSANEIRGRNRPQVDGVLAVVVPVGAADNODVRTSPSTQATKQGN	878
Qy	804	IYHOSDALDSQVYVEGFSNFOAFQSPQYTNVIAKNGDLFKSWGITQFENAPQVVSSE	863
Db	879	VYESNAALDSQVYVEGFSNFOAFQSPQYTNVIAKNGDLFKSWGITQFENAPQVVSSE	938
Qy	864	DGTFDLSVILNGYAFSDRYDLAMSNNKYKSGQDLANAIKGLQSGAGIKVLSDLVNPQLYN	923
Db	939	DGSFLDSVIQNGYAFSDRYDLAMSNNKYKSGQDLANAIKGLQSGAGIKVLSDLVNPQLYN	998
Qy	924	LPGKEVTVATRVNQYQAKSGATINKTPYVANTRSYG--DYQEOYCGKFLDLOKLYPLRF	982
Db	999	LPGDEVTVATRVNNYGETKGAIDHLSYAAKTRFTFGNDYQGYGAGFLDELKRLYPQIF	1058
Qy	983	STKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVYLSEG--NKYLLMADGKLFPTVL	1041
Db	1059	DRVQISTGKMTTDEKITKWSAKYNGTILDRGSEYVLKNGLNGYGTNGGKVSPLKV	1118
Qy	1042	NNTYQGPQVSANG-----FISKNGGIHYLDKXNQBEVKRNP--KEISGSWYTFPSDGMATG	1095
Db	1119	GSNQSTNGDNQNGDGSKFEKRLFSVRYRNNQYAKNAFIKDNNDGNVYFVFNDSRMAVG	1178
Qy	1096	KTKIGNDTYLPWPNKQKOLKEGVWYDGK--KAYYDDNGRTWTN-----KGFB	1141
Db	1179	EKTIDGKYFFLANGVQLDRYQNRGQVFFYDQNGVLSANGKQDPDNNNNTSGRNQ	1238
Qy	1142	FRVDGQDKWRYFNGDGTIAIGLVSLDNRITLYFDAYGVQKQTVTINGKSYTFDADQDGL	1201
Db	1239	FVQIGNVWAYDYGNGKRVTHQNINGQBELFFDNNNGVQVKGRTVNGGAIRYDANSSEM	1298
Qy	1202	VQTDNANPAQOAGKWLGDNQWGY-RKDGQLLTGEGTIDGQKVPFDQNGVQVKGTTAT	1260
Db	1299	ARNRFAETEP-----GVWAYFNNDGTAVKGSQNINGQDLYFDQNGRQVKGALA-	1346
Qy	1261	DASGVLRFYDRDQGHQVKGWSTSDNNWVYVNESGOVLTLQTLTIDGQTVYFDDKGIQAK	1320
Db	1347	NVDGNLRYDVNSG-ELYRNRPHEIDGSWYFDGNGNAVKGWNINGQNLFFDNNKGQIK	1405
Qy	1321	GKAVWDEGNLRYFDADSGNMLDRWKNDV-GNWFYFNRNG	1360
Db	1406	CHLV-RVNGVVRVYFDPNSGEMAVNRWVEVSPGWVYFDGEG	1445

F:1330-1351/Domain: cpl repeat homology <CP4>
F:1352-1371/Domain: cpl repeat homology <CP5>
F:1402-1420/Domain: cpl repeat homology <CP6>
F:1465-1484/Domain: cpl repeat homology <CP7>
F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 41.8%; Score 3023; DB 2; Length 1592;
Best Local Similarity 41.3%; Pred. No. 1.4e-145; Indels 292; Gaps 36;
Matches 669; Conservative 221; Mismatches 436;

QY 1 MEKNRYKLHKYKQWVAIGVTVTV-LSPFAGQVVAADTNNNDGTSVQVKNWVSDPKF 59
DB 1 MEKNRYFKWHYKKEKRWVLSVASATMLASALGASVASADTDFTASDSSQNT--VVTGD--- 55

QY 60 DAQAQNGQLAQAAMFAANQADQATQTSQVSPATDGRVNDQVTPAANQPAANVANQDV---- 115
DB 56 --QTTNNQATQTSIAA-----TATSEQSASTDAATDQ--ASAAEQTCQTASTDTAAQT 106

QY 116 ---ANPA-----TDAGALNRQSAADTSTDGKAVPQTSQDPGHLETVDGKTYV 160
DB 107 TTNANEAKWVPTENENQGTDEMLBAKNVATAESD--SIPSDLAKSNVKNQVDKYY 164

QY 161 DANGQRLKNYSWVIDGKTYIFDQGTG-----EAQTDLPKTCQANQDNPDPSVQANNQAYS 216
DB 165 DQGNVKNFPAVSGDKIYFD-ETGAYKDTSKVDADKSSAVSQNA-TIFAANREAYST 222

QY 217 EASSFETVDNYLTADSWYRPRKILKNQSGWQASSEGDLRPLMTWVPPAAATKAAVANFWA 276
DB 223 SAKNFEADVNYLTADSWYRPRKILKDKGTWTSKGDDFRPLLMAWVPTETKRNVTYVNN 282

QY 277 KEGLSGSY--RQNSANLDAANTQISALEKKIASEGTNWLRDKMSQFVKSONQWSTAS 334
DB 283 KVVGDITKYTASTSQADLTAAAEVLQVAREQKITSENNTKLUREAISAFVKTPQPMNGES 342

QY 335 ENETVYPNDHMOGGALLFPNSKD--TEHANSWRLNRRNPTQTGK-QKYFTTN----YA 388
DB 343 EK----PYDDHLONGALLFDNQTDLTPTQSNYRLNLRTPNTQTSLSRSFYFNPDPILG 398

QY 389 GYELLANDVDSNPVQAEQLNLHLHLYLMNWGDIYVNGDKOANFDGRVVDADVNDVNLQ 448
DB 399 GYDFELLANDVDSNPVQAEQLNLHLHLYLMNWGDIYVANDADANFDSIRVDAEDNVDAQ 458

QY 449 IQRDYKAKYCTDQNEKNAIDHLSILEAWSGNDYVYKQDNFSLSIDNQDSGMLKAFG 508
DB 459 ISSDYLKAAAGTIDKNNKANNVSVIWEAWSDNTPLYHDDGNLNMKNFLSLMSLSLA 518

QY 509 YASAYRGNLSNLATAGLKNRSANP--DSDPVNPVFIKRAHDSVQTRIAKIIREKLKGTNA 567
DB 519 KFTDVRSGNLPLHNSLVDREVDDREVETVPSYSPARAHDSVQDIIRDIIKAEI-NNPS 577

QY 568 DGLTNLTLDLNLKAFDIYNQDMNATDKVYYPNNLPMAYAMWLNQDKTTRVYVYGDMYTDN 627
DB 578 FGYSFTQBEIDQAFKIYNEIDLKSKDKYTHYNVPLSYTLTLNKGSIPTVYVYGDYMTDD 636

QY 628 GQYMATKTPFYNAETLLKGRIKYVAGGQAVSYKQDWSSGILTSVRYGKANSASDAGNT 687
DB 637 GQYMAKNTVYDAIESLLEKARMKYVAGGQAMQYQIGNGEILTSVRYGKALKQSDKGA 696

QY 688 ETRNSGALLINRNPFA--YRNLTLNMGAAHKSQAYRPLLLSTKDGATYLNDSVDV-S 745
DB 697 TTRTSGVGVVGNQPNFSLDGKVALNMGAAHNGEYRALMVSTKDGAVATYATDASAKA 756

QY 746 ROYKYTDSQGNISFSASELSQSVANAQVSGMIQVWPVGAADNQDVRTSPSTQATKGNITY 805
DB 757 GLVKRTDENGILYFLNDDLKGVANPQVSGFLQVWPVGAADQDQIRVAASDTASTDGKSL 816

QY 806 HQSDALDSQVIEGGSNFOAPQSPDYTNNAVIAKNGDLFKSWGITQFEMAPQVYSSSDG 865
DB 817 HQDAAMDSRWMEGGSNFOFSKTEEXYTNVVIANNVDFKVSNGITDFEMAPQVYSSSDG 876

QY 866 TFLDSVLNGYAFSDRYDLAMSKNNKYSKQDLANAIKGLQAGIKVLSLDVLPNQLYNLP 925
DB 877 QFQDSVQNGYAFTRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMYTFP 936

QY 926 GKEVVTATRVNOYQAKSGATINKTPYVANTRSYG-DYOEQYGGKFLDLOKLYPRLPST 984
DB 937 KQEVVTITRTDFKPIAGSQINHSLSLYVTDFKSSGDDYQAKYGGAFDELKEKYPFLFTK 996

QY 985 KQISTGKPIDPSKIVITNWSAKYFNGSNILGRGAKYVLS--GNKYLNILADGKLFPLTVLN 1042
DB 997 KQISTGQALDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSNKYFNVAVDTLFLPSSIL 1056

QY 1043 NTYQOPVSANGFTSXNGSIHYLDKNGQEVKNRPFKIBSGSWYFSDSGKMATGKTKIGND 1102
DB 1057 GKWVESGIRYDG---KGYIYNSSATGDQVKASFTTEAGNLIFYFGDKGVMYVTAOTINGA 1112

QY 1103 TYLFWPCKQLKEGVVWDGK-KAYYDDNGRTWNK-CFVEFRVDQDKWYKENGSGTIA 1160
DB 1113 NYFFLENGTALRNTIYTDAQNSHYANDGKRYENENGYYQGF---GND-WRYFK-DGNWA 1167

QY 1161 IGLVSLD----- 1167
DB 1168 VGLTVDGNVQYFDKDGVOAKDKIIVTDFGKRVYFDQHNGNAVNTFTIADKTHGWYILGK 1227

QY 1168 -----NRTLYPDAYGVYKQGTVTIN-GKSYTFDADQDGL----- 1201
DB 1228 DGVAVTGAQTVGKQKLYPEANGQOVKGFVTSDEGKLYFYDVSQDMWMTDTFIEDKAGNW 1287

QY 1202 ----- 1201
DB 1288 FYLGKQGAAVTGAQTIHQKLYFKANGQOVKGDIVKGTGDKIRYYDAKSGQSVFNKTVKA 1347

QY 1202 -----VOTDNANPAP-----OCQ-----AGMKLLGDNQWGYRKD 1230
DB 1348 ADGTYVYICNDGVAVDVSVVKGQTFKDGASGALRPNYLGQLVTSQGWYETANHDVYIQS 1407

QY 1231 GQLITGEQITDQKVFFQDNGVQVKGGTATDASGVLFYDRDQGHQ----- 1276
DB 1408 GKALTGEQITNGQHLFYKKDGHQVKGVLVTGTDGKRVYYDANSQDAFNKSVTVNGKTY 1467

QY 1277 -----VGKGYSTSDNNWYVYVNESQGVLTCL 1302
DB 1468 FGNDDTAQTAGNPKQGTFFKDGSDIRFYSMEQQLVTGSGMYNQAQOWLYV-KNGKVLTLG 1526

QY 1303 QTIDQGTVYFDDKIQAKGKAVMDENGLRYPFDADSGNMLDRWKNVQDNWYFYFNRNG 1360
DB 1527 QTVGSQRYVFBENGIOAKGKAVRISDGKIRIFDENSGSMITNQWKEVNGRYYYFGNDG 1584

RESULT 7
B31135
gtfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C:Accession: B31135; A33128
J:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: B31135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SHI>
A:Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN: submitted to the Protein Sequence Database, September 1990
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171,173-641,'N',643-1475 <SH2>
A:Cross-references: UNIPARC:UPI000017AC5E
A:Experimental source: strain GS-5
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>

F;1354-1373/Domain: cpl repeat homology <CP4>
F;1419-1438/Domain: cpl repeat homology <CP5>

Query Match 40.7%; Score 2945.5; DB 2; Length 1475;
Best Local Similarity 45.8%; Pred. No. 1.1e-141;
Matches 652; Conservative 225; Mismatches 425; Indels 121;

Qy	1	MEKRLRYKLHKVKKQWVAIGVTT--VTLSFLAGGOVVAADTN-----NDGTSVQVVK	51
Db	1	MDKKVYKLRKKKRWVVVSVASAVMTLTTLSGG-LVKADSNESKQSIDNSNTSV----	55
Qy	52	MPSPDKFDAQNGQOLAQAMFK---AANQADQT-----ATQSVSPATDGRVDNQVTPA	102
Db	56	-----VTANEESNVITEATSKQEAASSQTNHTVTTSSSTSVVNPKE--VVSNPYT--	104
Qy	103	ANQPAAN---VANQDVANPATDAGALNRQSAADTSDGKAVP-----QTSDDPGHLET	152
Db	105	VGETASNGBKLNQNTTVVDKTTSEAAANNISKQTTTEADTDVDDNSAANLQILEKLPNVKE	164
Qy	153	VDCKTYYVDANGORLKNYSMWIDGKTYFYFDGTGEAQTDLPKTKGOANQDNV---PDSYQA	209
Db	165	IDGKYYYDNNCKVTRNFTLIADGKILHFD-ETG-AYTD-TSIDTVNKDITVTRSNLYKK	221
Qy	210	NNOQVSNSEASSPETYVDNYITADSWTRPKKILKNGSQWQASSEGDLRPILMTWPPDAATKA	269
Db	222	YNQVYDRSAQSFEHVDHLYLTABSWTRPKYILKDGKTTQSTEKDPRLLMTWPPDQBTQR	281
Qy	270	AVANFAWEKGLTSGSY--RQNSANLDAATQNIQSAIEKKIASEGNTWNLKDKMSQFVKSQ	327
Db	282	QVYVYMAQLGINKTYDDTSSNQLQNLNIAAATYQAKIEAKITTLKNTDMLAQTTISAFVKIQ	341
Qy	328	NQWSTASENETVYPNQDHMQGGALLFSN-SKDTSEHANSDWRLLNRPNTFOTGKQ--KYFT	384
Db	342	SAWNSDSEK----PFDDHLQNGAVLYDNEGKLTYPYANSRYAILNRPNTPTNGKKDPRYTA	397
Qy	385	TW-YAGYELLANDVDSNPVQVQAEOLNHLVNLWNGDVIWNGDKDANFDGVRVDAVDVYN	443
Db	398	DNITGGEYFELLANDVDSNPVQVQAEOLNHLHFLNMFNGNIYANDPDPAFDSIRVDAVDNDV	457
Qy	444	ADLLQIORDYRKAKYGTQDNEKNAIDHLSILEAWSGNDNDYVKDQNNFSLSDNDORSQM	503
Db	458	ADLLQIAGDYLKAAAGIHKNDKAAANDHLSILEANSNDTPYLHDDGDMNMNDKLRLSL	517
Qy	504	LKAFGVSAYRGNLSNLATAGLKORS-ANPDSDPVNPVYFIRAHDSVQTRIAKITREKL	562
Db	518	LFSLAKPLNQRSGMNPNTNSLVNRTDQNAETAAPVSPYSFIRAHDSVQDLADIKAEL	577
Qy	563	GKTNADGLTNLTLDLNLKAFDINYQDMNATDKVYYPNNLPMAYAWMLONKDTVTRYVYGD	622
Db	578	-NPNVUGY-SFTMEIEKAFIYNKDLATEKKYTHYNTALSYALLTNNKSSVPRVYVYGD	635
Qy	623	MYTDNGQYMATKPPYNAIETLLKGRIKYVAGGOAVSYKQDWSGILTSVRYGKGANSAS	682
Db	636	MFTDDGQYMAHKTINYEAIETLLKARIKYVSGGQMRNQVGNSEIITSVRYGKGALKAT	695
Qy	683	DAGNETRNSGMALLINNRPNFR--AYRNLTLANWGAHKQOAYRPLLSTKCGIATVYLD	740
Db	696	DTGDRITTSYGAVYIEGNNPSELKASDRVVVNWGAHKQOAYRPLLSTTDMGKAYHSD	755
Qy	741	SDVDSRQYKYTDSQGNLSFSAESQSVANAQVSGMIQVWVPVGA--DNQD-789	
Db	756	QEA-AGLVRYTNDRGELIFTAAIDIKGYANPQVSYGLGVVVPVGAALIKMFALRLARPHQQ	814
Qy	790	VRTSPSTQNTKDGNIYHQSDALDSQVIYEGESNFQAFQAPDQDYTNVIAKNGDLPKSWG	849
Db	815	MAS-----VHQNALDSRWMEFEGSFNFQAFATKKEEYTNVVIANKVDKFAEWG	862
Qy	850	ITQFEMAPQYVSSDGTFLDLSVILNGVAFSDRYDLAMSKNNKYGSKODLANAIKGLQASG	909
Db	863	VTFDEMAPQYVSSDGSFLDSVIQNGYAFTRDYDLGISKENKYGTADDLVKAIKALHSGK	922
Qy	910	IKVLSDLVPNQLYNLPGKEVVTATRVNQYQAKSGATINTKTPYVANTRSYG-DYQSOYGG	968

Db	470	VDAVDNDVADMLQIYTYNFRYFYGVNKSEANALAHISVLEAWSLNDHNHYNDKTDGALAM	529
Qy	496	DNDORSGLMKAFGYA-----SAYRGNLSNLATAGLKORSANPD-----	533
Db	530	ENKQRLALL--FSLAKPIKERTPAVSPLYNNTFNTTORDEKTDWINKDGSKAYNEDGTVK	587
Qy	534	-----SDVPNVYFIRAHDSVQTRIAKIREKLGKTNADGLTNLTLDLNNKAF	582
Db	588	QSTIGKNEYGDASGNVYFIRAHNDNVQDIIABIIKKEI-NPKSDGFT--ITDAEMKQAF	645
Qy	583	DIYNQDNNAKTDKVYYPNNLPMAYAWMLQNKDVTTRVYVGYDMYTDNGOYMATKPTFYNAIE	642
Db	646	EIYNKMLSSDKKTYLLNNIPAAVAMVLQNMETITRVYIGDLYTDDGHYMETKSPFYDITV	705
Qy	643	TLKGRIKYVAGQAV--SY-----KODWS-----SGILTSVRYGKGANSADA--GNT	687
Db	706	NLMKSRIKYVSGGAQORSYMLPTDCKMDNSDVELYRTNEVTVSVRYGKDITMANTBEGSK	765
Qy	688	ETRNSCHALLINRPNPFRAYNLTLN--MGAAHKSQAYRPELLISTKDGIATYLNDSVDVS	745
Db	766	YSRTSGQVTLVANNPKLNLQDSAKLANMEMGKIHANQKYRALIVGTADGKNFTSDADAIA	825
Qy	746	RQY--KYTDSQCNLSFSASELSQSVANAQVSGMIQVWPVPGAADNOVRTSPSTQATKOGNI	804
Db	826	AGYVKETDSNGVLTFGANDIKGYETPDMSGFVAWVPVGASDNQDIRVAESTEAKKEGEL	885
Qy	805	-YHOSDALDSQVIYEGFSNFQAP--AQSPOQYTNNAVIAKNGDILPKSHGIIQTQFEMAPQYVS	861
Db	886	TLKATEAYDSQIIYEGFSNFQTTIPDGSDDPSVYTNRKIAENVDILFKSNGVTSFEMAPQVS	945
Qy	862	SEDCTFLDSVLNGYAFSDRYDLAWSKNNKYGSKODLANAIKGLQSAGIKVLSDLPVNQL	921
Db	946	ADDTGTFLDLSVIQNGYAFADRYDLAWSKNNKYGSKEDLRDAIKALHRACIQAIADWPDQI	1005
Qy	922	YNLPGEVVTATRVNOYGOAKSGATINKTPIYVANTRSYG--DYQFYQGGKFLDLQKLYPR	980
Db	1006	YQLPGKEVVTATRTDGAGRKIADAIIDHSLYLVANSKSSGKDYOQAKYGEBFLAELKAKYPE	1065
Qy	981	LFSTKQIOTSGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSE--GNKYLNLADGKFLP	1038
Db	1066	MFKVNMISTGKPIDDSVKLQKWKAEBYFNGTNVLERGVGIVLSDEATGKYFTVTKBEGNFIP	1125
Qy	1039	TVLNNTYGPQPVSAINGPISKNGGIHYLDKNGQOEYKRNFKBISGSWYFYSDGKMATGK--	1096
Db	1126	LQLT---GKEKV-ITGFSDDGKIGITYFTGSGTQAKSAFVTFNGNTYFYFDARGHVMTNSEY	1181
Qy	1097	TKIGNDYILFWPNCKQLKEGWYDCK--KAYIYDDNGRTWNTKGFVERV-----DGQD--	1148
Db	1182	SPNGKQYVFLPNGIMLSNAPYIDANGNTYLYNSKGQMYKG-GYTKFEDVSETDKDGKESK	1240
Qy	1149	--KWRVFNFGDGTIAIGLVSLDNRTLFPDAYGYQVKGQTVTINGKSYTFDADQGDVLQDN	1206
Db	1241	VVKERYFTNEGVMAGVTVIDGFTQYFGEBDGFOAKDLVTFKGTYYFYDAHTGNGIKDT-	1299
Qy	1207	ANPAPOQGAHWKLGDNQWGYRKDGQLLTGEBTIDGQKVPFQDQNGVOVKGGTATDASGVL	1266
Db	1300	-----J-----WRNNGWKYTFDANGVAATGAQVINGOKLYFNEBDGSQVKGWGVKNADGTY	1349
Qy	1267	RFYDRDQGHQVKGWYSTSDNNWVYNESGOVLTLGLQITDQGTVYPDDKGIQAQKGAAMD	1326
Db	1350	SKYKEGFELVTNEFFTGDGNVYIYAGANGKVTGQAVINGQHLFYFNADGSQVKGWGVKN	1409
Qy	1327	ENGLNRYFDADSGNMLDRWKQV--DGNWYFYNRNG	1360
Db	1410	ADGTYSKYNASTGBRLTNEPFTTGDNNWYIIGANG	1444

C;Accession: T31098
R.;Monchois, V.; Remaud-Simeon, M.; Monsan, P.; Willemot, R.M.
FEMS Microbiol. Lett. 159, 307-315, 1998
A;Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (D
A;Reference number: Z20981; MUID:98164374; PMID:9503626
A;Accession: T31098
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1508 <MON>
A;Cross-references: UNIPROT:O52224; UNIPARC:UPT00000BB69B; EMBL:AF030129; NID:g2766611;
A;Experimental source: strain NRRL B-1299
C;Genetics:
A;Gene: dsrB
C;Function:
A;Description: produces dextran composed only of alpha (1-6) glucosidic bonds
C;Keywords: glycosyltransferase; hexosyltransferase

```
Qy 675 GKANSADAGNTETRNHSGMALLNNRPN--FRAYRNLTLMNGAAHKSQAYRPLLLSTKD 732
Dy 767 GKGMVTSNADTRTGGIVIVSNKENLALKSGDVTLLHMGAAHKNQAFRLILGTTAD 826
Qy 733 GIATYLNDSVDSDROYKYTDSQGNLSPSASELSQVANAQVGMIOQWVPVGAADNODVRT 792
Dy 827 NLSYV----DNDNAPVKYTNDDQGLIIFDTEIYGVNRPQVSGFLAVWVPVGAOSHQDART 882
Qy 793 SPSTQATKQGNIIHQSDALDSQVIYEGFSNFQAFAPQSDQYTNNAVIAKNGDLFKSMGITQ 852
Dy 883 LSDDTAHDGKTFHNSNAALDSQVIYEGFSNFQAFATNEDYTNNAVIAKNGQLFKMGITS 942
Qy 853 FENAPQVYSSDGTFLDSVLNGYAFSDRDVLAMSKNKKYGSQDLANAKGLQSGAGIKV 912
Dy 943 FOLAPQVRSSTDSFLDSIIQNGYAFDTRDGLYGTGTGTVDQLDADAIKALHANGIOA 1002
Qy 913 LSDLPVNLPGKEVVTATRVNOYGOAKSGATINKTPVANTRSVGDYQEOYGGKFLD 972
Dy 1003 IADWVPOIYNLPQELATVTRNSYGDQDNTSDIDQSLVIOGRGGKYQAQYGGAFLS 1062
Qy 973 DLQKLYPRLFTKQISTGKPIDSPSVKITNWSAKYFNGSNILGRGAKYVLSEGN-----KY 1027
Dy 1063 DIQKYPALFETKQISTGLPMDPSQKITESGKYFNGSNIOGKGAGYVLKSDGTDQYKYV 1122
Qy 1028 LNLADGKFLPTVLNNTYGOQVQVANGFISKNGGIIHYLDKNGQEVKNRP-KEISGWSYVF 1086
Dy 1123 TSNNNRDFLPKQLTD-----DLSETGFVRDNGWYVYTLGVLARNTFIQDNGNYYVF 1177
Qy 1087 DSPGKWATGKTLGNDTYLFWPBGKQLKEGVWY--DGKAYYYDDNGRTWTKGFEPRV 1144
Dy 1178 DSTGHLVTGPNINHHYFLPGLNGIELVQSLQADG-STIYFPDQGR-----QVFNQYIT 1232
Qy 1145 DGQDKWRYFNGDGT-IAIGLVSLDNRTLYPDAYGYOVKGQTVT-INGKSYTFDADOGDLV 1202
Dy 1233 DQGTAYYFQNDGTWYTSGFTEIDCHKQYFYKNGTVQVKGQFVSDTDGHVYLEAGNGN-V 1291
Qy 1203 QTDNANPAPQOGAKWLLGNQWYKRDGOLLTGEBOTIDGQYVFPDNGVQVKG-----1257
Dy 1292 ATORFAQNSQGO--WFLNG-----DGIALTGLQTINGVQVYFVADGHQSKGDFITIQ 1342
Qy 1258 -----TATDASGVLYRFDYR-----DOGHQVKGWSTSDNWNVYNESQVLTGLQ 1303
Dy 1343 NHVLYTNPLTGATIGMQQIGDKIFVFDNTGNMLTNQYQYOTLQDQWHLSTQGPADTGLV 1402
Qy 1304 TIDGQTVYFPDQKGIQAKGKAVMDENGRLRYF-DADSGNMLRDRKWNKVDGWNWYFERN 1359
Dy 1403 NINGNLKYFQANGRVKQGVPTDPTNVSYMATDGSVAVDNYFTYQGWYLTDSN 1459

RESULT 12
JC5473
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides
C:Species: Leuconostoc mesenteroides
C:Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
C:Accession: JC5473
R:Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.
Gene 182, 23-32, 1996
A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconoc
A:Reference number: JC5473; MUID:97136686; PMID:8982063
A:Accession: JC5473
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1290 <MON>
A:Cross-references: UNIPROT:Q48756; UNIPARC:UPI000017ABDF; GB:U38181
C:Comment: This enzyme catalyzes the transfer of D-Glucopyranosyl units from sucrose on
C:Genetics:
A:Gene: dextr
C:Keywords: glycosyltransferase; hexosyltransferase
F:78-870/Domain: catalytic #status predicted <CAT>
F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 35.9%; Score 2593.5; DB 2; Length 1290;
Best Local similarity 42.2%; Pred. No. 6.8e-124;
```

```
Matches 543; Conservative 228; Mismatches 395; Indels 121; Gaps 24;
Qy 149 HLETVGKTYVDANGORLKNYSWIDGKTYYPFDGOTGEAQDTOLPKTQGANQNVDPDSYQ 208
Dy 3 NVKVDGKYPYGDGQPKQNFTHIDGKPYFPDKDTGALSNN-----DKQVSELFS 55
Qy 209 ANNQ-----AYSNEASSFETVDNYLTADSWYRPRKILKNQSQWASSEGDRLPILMTWMPDA 265
Dy 56 IGKHNNAVNTSNDNFQLEGHLTASSWYRPRKIDILKNKRWAPSTVTPRPLLMAMWMDK 115
Qy 266 ATCAAYANFWAKBGLISGSY-----RONSANLDAATQIQSAIEKKIASEGNTWLRDKMS 321
Dy 116 STQVTVLNTYKQGLLSGTHHFSFDNENMRTLTAAMQAOVNIIEKKIQLGQNTWLTAMT 175
Qy 322 QFVKSQNSIASENETVVPNOHMOGALLFNSKDTSEHANSDWLLNRPNFTQTK-- 379
Dy 176 QYIDAQPNWIDSEAK-----GDDHLOGGALLYTNSDMSPKANSDYRKLSTPKNQKGOIA 231
Qy 380 QKYFTTNYAGYELLANDVDNSNPVVOAEQLNHLHYLMNMGDIWVGDKDANFQGVVRDVA 439
Dy 232 DKY---KOGGFELLANDVDNSNPVVOAEQLNHLHYLMNMGDIWVGDKDANFQGVVRDVA 288
Qy 440 DNVNADLLQIQRDYKAKYGTQDNEKVAIDHLSILSEAWGNDNDYKQDN-----489
Dy 289 DNVADLLQIAGEYAKAAYGVD---KMTREINIIYQFWKTGE---MKIQTMSKHMATSKL 342
Qy 490 -----NFSLSIDNDQSRGMLKAFGVSAYRGNLSNLTAGIKNRSANPDSDPVNP 539
Dy 343 SMDPPLHLAIKYALNPNNDKRSGLPTREHSL-----VKRITDDKENVAQPN 389
Qy 540 YVFIRAHDSQVTRIAKIIREKLGKTNADGLNLTLDLNLKAFDIYNQDMNATDKVYYPN 599
Dy 390 YSFIRAHDSQVTRIADIIKINPASTGLDSTVTLDOIQAQFDIYNADDELKADKYVTPY 449
Qy 600 NLPMAVAMWLNQKDTVTRVYGYDMYTDNQYMATKTPFNATETLKGRIKYVAGQAA-- 657
Dy 450 NIPASVALLPLPKNDTIPRVYGYDMFTDQGYMAKOSPPYCAIDALLKARIKYAAGQOTMK 509
Qy 658 VSYKODWSSGILTSVRYGKGNASADAGNTETRNHSGMALLNNRPNFR--YRNLTLNMG 715
Dy 510 MYFPDEQ--VMISVRYGKGNATASDGNQETRYQGILVNVNRPDLKLSDDKDEVKDDMG 568
Qy 716 AAKHSQAYRPLLLSTKQGIATYLNDSVDSDROYKYTDSQGNLSPSASELSQVANAQVSGM 775
Dy 569 AAKHQDYRPLVLTTSKGLKYVTDANAPVVR---TDANGQLTFKADMYGVNDPQVSGY 625
Qy 776 IQWVPVGAADNQVTRTSPSTQATKQGNIIHQSDALDSQVIYEGFSNFQAFAPQSDQYTN 835
Dy 626 IAAWVPVGAENQDARTKSETTQSTDGSVYHNSNAALDSQVIYEGFSNFQDFTPTTDEFTN 685
Qy 836 AVIAKNGDLFKSGWITQPEMAPQYVSSDGTFLDSVLNGYAFSDRDVLAMSKNKKYGSK 895
Dy 686 IKIAQNVNLFKDWGITSFEMAPQYRASSDKSLDAIVQNGYAFPTDRYDIYGNTPTKYGT 745
Qy 896 ODLANAIKGLQSGAGIKVLSDLVPLNQYLNLPKGVEVVTATRVNOYGOAKSGATINKTPYVAN 955
Dy 746 DNLLDALRALHQGGIQAINDWVDPQIYNLPDSQLVTAIRTDGSGDHTYGSVIDHTLYASK 805
Qy 956 TRSYGDYQEOYGGKFLDQLKLYPLFSTKQISTGKPIDSPSVKITNWSAKYFNGSNILGR 1015
Dy 805 TVAGGIYQQYGGAFLEQLKTYQPLFQKQKISTQDPMNPDIQIKSWEAKYFNGSNIQGR 865
Qy 1016 GAKYVYLSE--GNKYLNLADGKLPLTVLNTNTYGOQVQVANGFISKNGGIIHYLDKNGQEVK 1073
Dy 866 GAWYVLKDWGTOQYFNVSDAOTFLPKQLLG-----EKAKTGFVTRGKETSFYSTGYQAK 920
Qy 1074 NRFKEISGSWYFSDSGMKTGKTKIGNDITYLFPNGKQKKEGVWYDGKAYYDDNGRT 1133
Dy 921 SAFICDNGWYVFDKGRKAVNGVINGINYFLPNGIELQDAYLVHDGMYYYNNIGKQ 980
Qy 1134 WTKNGFVFRVDQDKWRYFNGDGTIAIGLVSLDNR-----TLYFDAYGYQVKGQTVT-I 1187
Dy 981 LHN-----TVYQDKQKNFHYFFEDGHMAQGIIVTIIQSDGTFVTFQYFDENGKQKGVAVKGS 1036
```

QY	1188	NGKSYTFDADQGLV-----QTDN-----ANPAPQGGAG	1216
DB	1037	DGHLHYFDGASGNLPSKRWGLADGSLWYDEKGNVAVTGKQTINNQTVYFNDGGRQIKNN	1096
QY	1217	WKLGLDGNQGY-RKDGQLLTGQTIDGQVFFQDNGVQVKGGTATDASGVLFYDRDOGH	1275
DB	1097	FRELADGSLWYLNKKGVAVTGSEIINGQTLFPGDGRQPKGTHINATGESYYPDDSGN	1156
QY	1276	QVKGWYSTSDNNWVYNESGOVLTGLQTIDGQTVYFDDKGIQAKGKAVMDENGLRYPD	1335
DB	1157	MITDRERVDNQWAFYGDGVAVTGDRIIKGNLYFNQNGIQMKGH-LRLENGIMRYD	1215
QY	1336	ASGNNMLRDRWKNV-DGNWYFNRNGL	1361
DB	1216	AUTGELVRNRFVLLSDGSLWVYFQDGV	1242
RESULT 13			
B97033			
uncharacterized protein, related to enterotoxins of other Clostridiales [imported] - Cl			
C:Species: Clostridium acetobutylicum			
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004			
C:Accession: B97033			
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,			
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.			
J. Bacteriol. 183, 4823-4838, 2001			
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl			
A:Reference number: A96900; MUID:21359325; PMID:21359325			
A:Accession: B97033			
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-2817 <KUR>			
A:Cross-references: UNIPROT:Q97K42; UNIPARC:UPI00000CA0A0; GB:AE001437; PIDN:AAK79053.1;			
A:Experimental source: Clostridium acetobutylicum ATCC824			
C:Genetics:			
A:Gene: CAC1079			
Query Match 7.5%; Score 543; DB 2; Length 2817;			
Best Local Similarity 22.7%; Pred. No. 4.2e-19;			
Matches 357; Conservative 190; Mismatches 542; Indels 486; Gaps 89;			
QY	134	STDGKAVPOTSDOPHLEIVDGTYYVDANGORLKNYSVVIDGKTYFF-----	181
DB	451	ANDGKLLT-----GLQVINGSYFYDTNGIRLVSRTITIDGKYFNDQGILTDNWIN	503
QY	182	DG-----QTGEAQT-----LPKTQANQDNVP-----DSY	207
DB	504	YDGKYFYISGVKQTGLQIDNGYFYFSDSGIMQTGLQIDGKTYFPGDNGIRQIGWITY	563
QY	208	QANNQAYSNEASGFET---VDNYLTADSWYRPR-----KILK-----NQSQWQASSE	251
DB	564	Q-NKKYFNSDGSMTDLKIYSYSTSPYNVHYQYGFNDGKLLTGLQTIKNTYFYFDSN	622
QY	252	GDLRPLMTWPDAAATKAAAYFWAKEGLISGYSRONSANLDAATONIOSATEKKIASE-	310
DB	623	G-----ISQMGWNIDKDFYN---SNSIMTENWVINDKYYFYINNVRKQTGFQYINGKY	675
QY	311	-----GNTNMLRD---KMSQFVKSQ-----NOMSIASENE	337
DB	676	YFDPDGIQMTGQTISGNTYYLDDNGVKTGQVHTIKGKYDFDNGVGMVNTW-VFDNDK	734
QY	338	TYVPNDHMQGGALLFSNS-----KOTEHANSOWRLN-----RNPTFQT-----GKQ	380
DB	735	TYVYINGNNQMGTAISNNHYFGDDNGIMQTGQWRINGRTYFYFDDNNGAAKTGLVTEGKT	794
QY	381	KYFTTNYAGYE---LLAND---VDNSNPVQABQLNH---LHLYMNGDIIWMDK---DA	429
DB	795	YYPNTTYAYLDTGFIYFNNNYFPLDN-NGVVRTGMINYNNRYLSDSTGVRVTGFTQIDG	853
QY	430	N---PD--GVRVDAVDNVDLLOIQORDYKAKYGTQDNEKNAIDHLSILEAW-----SGN	480
DB	854	NKXYFSDSSGAMCTSFITVNGN-----TYGF---SKDGI-----MLTGQHTILSN	895

QY	481	DNDYKQDNFSLSDNDORS-----MLKAFGYASA---YRGNLSNL	520
DB	896	YSSY-----NIYFNSDGSQAQGFYTLGKTYFEPNYGYMLLGYNYINGKYYFNDGVI	951
QY	521	ATAGLKNRANPDSDPVENVV--FIRAHDSVQVTRIAKIIREKLKGTNAD-----	568
DB	952	QTGWTDSSKYYLDPGAAVTGTFQNIINGDKYFNSSGIMQTGLVYVNPDYGFDDNGHI	1011
QY	569	--GLTNLT-----LDDLKA-----FDIYNQDMNATDKVY--PNNLPMAY	605
DB	1012	LTGHSINGIYIYFDPSTCAQKGFVYTLGKTYFTNTMYTGFPVANNLLYFDPNEGVMOT	1071
QY	606	AWMLQNKDVTTRVY-----GMYTDNGOY-MATKTPFNALTELLKGRIKYVAGQ	656
DB	1072	GMINTNSN---RYFVSATGASVTGFTIDGNKICYFDSNGAIYDVVTI-----NGS	1119
QY	657	AVSYKQDNSSGIL---TSVRYKGKANSASDAGNTETRNS-GMALLINNRPNFRAYRLT	711
DB	1120	TYGFNTD---GIMLTGWTIRINRGYSSYF---NTYFNSDGT-----KTGFFYLNKT	1168
QY	712	LNMGAAHKSQAYRPLLLSTKDGIAIYLNDSVDVDSRQ-YKYTDSQGNLSFSASELSVANA	770
DB	1169	Y-----YFNPSDGERMLQGYQYIN--GNHYVFAD-----	1195
QY	771	QVSGMIQV-WV-----PVGADNODVNTSPSTQATKD-----GNIYH	806
DB	1196	--GTMQTGWITNGSSKTYLDPGSAAVT-GLQTINGNKYCFDSNGILQHNGIFYIGNTY	1251
QY	807	QSDALDSQVIVYEGFSNFQAF-----QSPDQYTNNAVIAKNGD--LPKSWGITQFEMAPQ--Y	859
DB	1252	GSD--NNGIMLTGLQLINGILYCFNSDGSVKTGLVYLGKTYFDSYSVSGGQINNNY	1309
QY	860	VSSDGTFLDSVILNGYAFSDRYDLAMS-----KNKK-----YGSKODLANA---	901
DB	1310	YFGNDGTWQTGWVNYGY--RYLNDSGIKVYTGQTTIDGNKYFYDYGYAKTGIVNIDGN	1366
QY	902	IKGLOSAGIKVLS-DLVPNLNLPGKEVVTATRV-----NOYGOAKSGA-TINKT	950
DB	1367	YYGFNSGVMLTGWOHINGSTYFNSNGIANTGTYLGKTYFDSYSGRMQIGSTINGT	1426
QY	951	PV-----ANTRSYGDYQEOYGGKFLDLOKLYPRLFSTQISTGKPIDPSV	997
DB	1427	SYFYANGVMKTSSTDSPNTLAVGWVDSY-----YYQYVYLNAAAGTK--LTG-----L	1471
QY	998	KITNWSAKYFNGSNILGRGAKYVLSSEKNKYLNLADGKLF--LPTVLNNY--GQPOVSAN	1053
DB	1472	QTIDGNTYFDSNGIMQTG--IITINGNRGFGVNGVMLYGLQFINNNYYSNGYSISQT	1529
QY	1054	GFISKNGGIHYLDKNGQEVKRRFKESGWSYFSDSGMKATGKTIGNDTYL--PMPNGK	1111
DB	1530	GFVTLISGNTYFDSYG-EMRIGLTYINNYYFNSKGMETGWI-----SYLRYANPNGI	1583
QY	1112	QLKEGVWVDGKAYYYDDNGRTWTKGFEPRVDGDKWRYFNGDGTIAIGLVSLDNRTL	1171
DB	1584	LUTGQTING-KTYFNSDGSLLYDQYINGSYFQDK-----NGVMYLGLOTICGNTY	1636
QY	1172	YFDAYGYQVKGQTVTITNGSKSYTFDADQGLVQTDNANP-----APQG--QAGMKLLGDNQ	1224
DB	1637	YLNNGISQSG-FITLNGKTYFYDSYGMRTGQINNNNYFFGDNGLTQGTWISQDNLR	1695
QY	1225	WGYRKDGQLLTGEQITIDGQKVF-----QDNQVQVKGGT--ATDASGVL-----RF	1268
DB	1696	YYANSSGVCLTGLQITIDGKYYFNSYARMETGLVYINNYYGDFNDGDTLLYSHWHNINGRM	1755
QY	1269	Y-----DRDQG-----HOVGKWTYSTDD	1287
DB	1756	YCFNTDGTWKTGWINYLGRSCYLDSSQGLSTGLLTIGHNIYYFSDSYSKMTGW-TSSG	1814
QY	1288	NWVYVNESQVLTGLQITIDGQTVYFDDKGIQAKGKAVMDENGLRIFYDADSGNMLDRWK	1347
DB	1815	SKYFYFNESGIMLTGFTIDGNTYFYDSYGNSTTGTTR--SINGNCYGFN-DDGIMLTG-WQ	1870


```
Db 2035 GFEYFAPANTYNNIEGQAIYVQSKFLTLNGKKYFDNNKSAVTGWQTID-----SK 2086
QY 1005 KYFNGSNILGRGAKYVLSGNYK-----LNLA-----DGKFL-----PTVL 1041
Db 2087 KYFNTNTAAEATGWTIDGKKYFNTNTAAEATGWTIDGKKYFNTNTAIASTGYTII 2146
QY 1042 N-----NTYGOPOVSANGFISKNGIHYLDKN-----GOEV--KRFKEISGSWYF 1086
Db 2147 NGKHFFYNTDGMQIGV--FKGPGFYEYFAPANTDANNIEGQAIIYQNEFLNGKKYF 2204
QY 1087 DSDGKMATKTKIGNDYLFMPNGKQKBEVWYDGKAYYDDNGRTWTKGFVFRVDG 1146
Db 2205 GSDSKAVTGWRIINNKYYFNNA-----IAAHLCTIN 2239
QY 1147 QDKWRYFNGDGTIAIGLSLDNRTLYFDA-----YGYQVK 1181
Db 2240 NDKY-YFSYDGLQNGYITERNPFYFDANNESKMVTGVFKGPGFYEYFAPANTENNIE 2298
QY 1182 GQTV-----TINGKSYTFDADQGLV--QT-----DNANPAQOAGWKLJGDQW 1225
Db 2299 GQAIYVQNKFLTLNGKKYFNDKSKAVTGWQIDGKKYFNLTAA- EAATGWQIDGKKY 2357
QY 1226 GYR-XDGQLLTGEQTIDGQKVFQDNGVQVKGGTATDASGVLFYDRDQGHQV----- 1278
Db 2358 YFNLTAAEATGWTIDGKKYFNTNTFIASGT-YTSINGKHFFYNTDGMQIGYFKGPN 2416
QY 1279 -----KGYSTSDNNVYNESGGVLTGLQTDGQTVYFD-DK 1315
Db 2417 GFEYFAPANTDANNIEGQAIIYQNKFLTLNGKKYFGSDSKAVTGLRTIDGKKYFNTNT 2476
QY 1316 GLOAKGKAWDE-NGNLYRFDADSGNMLDRKNVGNWYFNRLG 1361
Db 2477 AVAVTG---WQTINGKKYFNTNT-SIASTGYTIIISGRKHFFYNTDGI 2519

RESULT 15
140884
cytotoxin L - Clostridium sordellii
C:Species: Clostridium sordellii
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
C:Accession: I40884
R:Green, G.A.; Schue, V.; Monteil, H.
Gene 161, 57-61, 1995
A:Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium sordellii
A:Reference number: I40884; PMID:95369733; PMID:7642137
A:Accession: I40884
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2364 <RES>
A:Cross-references: UNIPROT:Q46342; UNIPARC:UPI00000802BE; EMBL:X82638; NID:g1000694; PIR:
C:Keywords: cytotoxin

Query Match 5.0%; Score 358.5; DB 2; Length 2364;
Best Local Similarity 20.3%; Pred. No. 7.6e-10;
Matches 233; Conservative 157; Mismatches 368; Indels 389; Gaps 56;

QY 397 DVDN-----SNPVQBAQLNHLHYLMNWGD---IVMGDKDANFDG-----VRV 436
Db 1347 DVDNVQKNTTIESIQGELIENILSKLNIENKILNNHTINPFYGDINESNRFISLTF 1406
QY 437 DAVDNVADLLQIQDYKAKY-----GTDQNEKNAIDHLSILEANSNDN 482
Db 1407 SILEDIN---IITIEDLVSKSYKILLSGNCMKLIENSDDIQK--IDHI-----GFNGEHQ 1457
QY 483 -----DYVKDQNFSLINDORSGLMKAFGVASRYGNLSNLTATAGLNKRSA-----NP 532
Db 1458 KIPIYSIDNEYKNGFYDYSKEG-----LFTAEFNSIIRNIYMP 1500
QY 533 DSDPVPNYVFIIRAHSEVQTRIAKIIREKLGKTNADGLTNLTDLDLANKAFDIYNQDMNAT 592
Db 1501 DS-----NNLFYSSKDLKDRI-----INKG-----DVKLL 1527
QY 593 DKVYVFNPLPMAYAWMLQNDVTRVYVYGDYTDNGQYMATKTPFPNAIETLLKGRIKV 652
```

```
Db 1528 IGNFKDDMKVLSFTIEDTWTIKL-----NGVYLDE-----NGVAQIL----- 1566
QY 653 AGQAVSYKQDWSGILTSVRYGKANSASDAGNTTETNSGMALL--INNRPNFRAYRNL 710
Db 1567 -----KFMNAKSAINTS--NSLMNFLESINIKNIP--YNNL 1599
QY 711 TLMGAHKSQAYRPLLLSTKDGIATY--LNDSDVDSRQY-----KYTDSOGNLSF 759
Db 1600 DPNTEFILDNTN-----FIISGNSIQGFELICDKDKNIQPYFINFKIETSITLYYGN--- 1652
QY 760 SASLQSVANAQVSGMIQVWPVPGAADNQDVRTSPSTQATKDGNIVHQSDALDSDSVIYEG 819
Db 1653 -----RQNLIVEPSYHLDDSGNI-----SSIVI--- 1675
QY 820 FSNF-QAFAQSPDQYTNNAVIAKNGDLFKSWGITQFEMAPQVYSSE-----DGFFLDSVI-L 873
Db 1676 --NFSQKLYGIDRYVNVKVIITAPNLYTDEINITPV-YKPNVICPEVILTDANYINEKINV 1732
QY 874 NYAFESDRY-----DLAMSKNKYGSQDL-----ANAIKQLQSAGIKVLSDLVFNQI- 921
Db 1733 NINDLSIRYVWNDGSDLLILIANSEEDNQPVKIRFVNVEK-----SDTAADKLUS 1782
QY 922 YNLPGEVVTATRVNOYGOAKSGATINKTPVYVANTRSYGDYQEOYGGKFLDQLQKLYPRL 981
Db 1783 FNFSDKQDVSVKI-----ISTFSLAAYSDFDY--EFLVSLDN-DYFYINS 1828
QY 982 FSTKQISTGPKIPSVKJITNNSAKYFN--GSNILGRGAKYVLSGNYK----- 1028
Db 1829 F-----GNMVSGLIYI-NDSLYYFKPKKNLI--TGFTTIDKNKYVDFPTKSGAASI 1877
QY 1029 --NLADGKLFPLTVLNNTYGOPOVSANGFISKNGGIHY-----LDKN--GOEVKNRPK- 1077
Db 1878 GEITIDGKDY----YFNKQGLQV--GVINTSDGLKYFAPAGTLDENLEGESVNFIGKL 1930
QY 1078 EISGSWYVFDSDGKMATCKTKIGNDYLFPMPNGKOLKEGVWYDGGKAYYDDNGRTWTK 1137
Db 1931 NIDGKIYFEDNYRAAEVWKLDDDETYFNPXTGEALQHLQIGDNKYFPDDNGIMQT-- 1988
QY 1138 GFVEPRVDGQKWRYFNGDGTIAIGLSLDNRTLYFDAYGVQVKG----- 1182
Db 1989 GFITI-----NDKVFYFNNDGVWQVGYIEVNGKYFYFGKNGERQLGVFNTPDGFKFGPKD 2044
QY 1183 -----QTVINGKSYTFD-----ADQGLVQ-TDNANPAQG--- 1213
Db 2045 DDLGTEGELTYNGILNFGNKIYFFDISNTAVWGMGTLDGSGTYFFDNNRAEACIGLTV 2104
QY 1214 -----QAGWKLLGDNOMGYRKDGQLLTGEQTDGQKVFQDNGVQVKG--- 1256
Db 2105 INDKKYFDDNGIRQLGFIITINDNIFYSESQKIELGYQNINGNYFYIDESGLVLIGVPD 2164
QY 1257 -----GTATDASGVLR-----FYDRDQGHQVQKGYSTSDNNVYV 1293
Db 2165 TPDGKYFAPLNTVNDNIYGOAVKYSGLVVRVNEDEVYFGETYKTIETGWIENETDKYFDP 2224
QY 1294 ESGQVLTGLQTDGQTVYFDDKGIQAKGKAVDENGNLRYFDADSGNMLRDRKKNVDGNW 1353
Db 2225 ETKKAYKGINVVDDIKKYFDENGIMRTGLISFENN--YYFNE--GKMQFGYLNLIKDKM 2280
QY 1354 YVFNRRNG 1360
Db 2281 FYFGKDG 2287
```

Search completed: February 11, 2006, 19:42:44
Job time : 40.1742 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:57:35 ; Search time 184.185 Seconds
(without alignments)
5228.676 Million cell updates/sec

Title: US-10-797-821-39

Perfect score: 7230

Sequence: 1 MEKNLYKLHKVKQKQWVIG.....WKNVDGNWYFNRNGLATRW 1365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2165443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7230	100.0	1365	1 GTFS STRDO	P29336 streptococ
2	5741.5	79.4	1338	2 Q9WJ4 9STRE	Q9WJ4 streptococ
3	3172	43.9	1462	1 GTFD STRMU	P49331 streptococ
4	3102.5	42.9	1577	2 Q54178 STRGN	Q54178 streptococ
5	3088	42.7	1575	2 Q9LCH3 STROR	Q9LCH3 streptococ
6	3080.5	42.6	1554	2 Q8KZL5 9STRE	Q8KZL5 streptococ
7	3079.5	42.6	1449	2 Q55264 STRSL	Q55264 streptococ
8	3076.5	42.6	1449	2 Q68542 STRSL	Q68542 streptococ
9	3052.5	42.2	1597	1 GTF1 STRDO	P11001 streptococ
10	3031	41.9	1590	2 Q59583 9STRE	Q59583 streptococ
11	3026.5	41.9	1577	2 Q55265 STRSL	Q55265 streptococ
12	3023	41.8	1592	1 GTF2 STRDO	P27470 streptococ
13	3020	41.8	1455	1 GTFC STRMU	P13470 streptococ
14	3018	41.7	1590	2 Q55263 9STRE	Q55263 streptococ
15	2992.5	41.4	1512	2 Q9WJ5 9STRE	Q9WJ5 streptococ
16	2992	41.4	1476	1 GTFB STRMU	P08987 streptococ
17	2975.5	41.2	1506	2 Q56CX8 9STRE	Q56CX8 streptococ
18	2942.5	40.7	1454	2 Q69A94 LEUME	Q69A94 leuconostoc
19	2891.5	40.0	2835	2 Q8G9Q2 LEUME	Q8G9Q2 leuconostoc
20	2863.5	39.6	1599	2 Q00599 STRSL	Q00599 streptococ
21	2804	38.8	1518	2 Q00600 STRSL	Q00600 streptococ
22	2702	37.4	1508	2 Q52224 LEUME	Q52224 leuconostoc
23	2692	37.2	1508	2 Q9EZHS LEUME	Q9EZHS leuconostoc
24	2687.5	37.2	1561	2 Q5SBM8 9LACO	Q5SBM8 lactobacill
25	2656.5	36.7	1477	2 Q9L466 LEUME	Q9L466 leuconostoc
26	2639.5	36.5	1330	2 Q84CN4 LEUME	Q84CN4 leuconostoc
27	2638	36.5	1527	2 Q8KEE1 LEUME	Q8KEE1 leuconostoc
28	2634	36.4	1522	2 Q6TXV4 LEUME	Q6TXV4 leuconostoc
29	2616	36.2	1527	2 Q9ZAR4 LEUME	Q9ZAR4 leuconostoc
30	2592.5	35.9	1290	2 Q48756 LEUME	Q48756 leuconostoc
31	2467	34.1	1463	2 Q5SBM6 LACFE	Q5SBM6 lactobacill

32 2399 33.2 1595 2 Q5SBM3 LACSK
33 2247.5 31.1 1772 2 Q5SBM3 LACRE
34 2227 30.8 2057 2 Q9RE05 LEUME
35 2193.5 30.3 1781 2 Q5SBL9 LACRE
36 2167 30.0 1772 2 Q5SBM0 LACRE
37 2148 29.7 1781 2 Q4UCS4 LACRE
38 2146 29.7 1781 2 Q9LJC7 LEUME
39 2065.5 28.6 1016 2 Q5SBM0 LACRE
40 1344 18.6 1619 2 Q5SBM0 LACRE
41 1320 18.3 1231 2 Q5SBM1 LACRE
42 1231 17.0 591 2 Q8VUH3 STRMU
43 923 12.8 522 2 Q8VV10 STRSA
44 670.5 9.3 374 2 Q6ZX19 9LACO
45 543 7.5 2817 2 Q97K42 CLOAB

ALIGNMENTS

RESULT 1

GTFS STRDO STANDARD; PRT; 1365 AA.
AC P29336;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfS;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE (GENOMIC DNA).
RX STRAIN=MFE28;
RC MEDLINE=90316665; PubMed=2142479;
RA Gilmore K.S., Russell R.R., Ferretti J.J.;
RT "Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltransferase that synthesizes soluble glucans.";
RL Infect. Immun. 58:2452-2458(1990).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -!- ENZYME REGULATION: Glucan synthesis by GTF-S is independent of primer glucan unlike GTF-I.
CC -!- MISCELLANEOUS: Synthesizes water-soluble glucans (alpha 1,6-glucose).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 10 cell wall binding repeats.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.
CC EMBL; M30943; AAA26898.1; -; Genomic_DNA.
CC HSP; P06653; 1H8G.
CC InterPro; IPR002479; Cell_wall_bd_put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 2.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
CC SIGNAL 1 36
CC Or 37 (Potential)
CC CHAIN 37 1365
CC Glucosyltransferase-S.
CC REPEAT 157 177
CC Cell wall binding 1.
CC REPEAT 178 197
CC Cell wall binding 2.
CC REPEAT 1062 1082
CC Cell wall binding 3.
CC REPEAT 1083 1102
CC Cell wall binding 4.

FT	REPEAT	1150	1169	Cell wall binding 5.
FT	REPEAT	1170	1190	Cell wall binding 6.
FT	REPEAT	1225	1243	Cell wall binding 7.
FT	REPEAT	1289	1308	Cell wall binding 8.
FT	REPEAT	1309	1328	Cell wall binding 9.
FT	REPEAT	1331	1352	Cell wall binding 10.
FT	REGION	198	1061	Catalytic (approximate).
SQ	SEQUENCE	1365 AA; 151591 MW; 167296B5A2B8C476 CRC64;		
Query Match				
Best Local Similarity		100.0%;	Score 7230; DB 1; Length 1365;	
Matches 1365; Conservative		0; Mismatches	0; Indels	0; Gaps
Qy	1	MEKNLRYLKHKVKKQWVAIGVTTVTLPSFLAGQVVAAADTNNDGTSVQVKNKMPSPDPKFD	60	
Db	1	MEKNLRYLKHKVKKQWVAIGVTTVTLPSFLAGQVVAAADTNNDGTSVQVKNKMPSPDKFD	60	
Qy	61	AAQNGQLAAQAFKAAQADQTATSVQSPATDGRVDNQVTPAANQPAANVANQDVANPAT	120	
Db	61	AAQNGQLAAQAFKAAQADQTATSVQSPATDGRVDNQVTPAANQPAANVANQDVANPAT	120	
Qy	121	DAGALNROSAADTSDGKAVPQTSPOQHLETFVDGKTYVVDANGORLKNYSWVIDGKTY	180	
Db	121	DAGALNROSAADTSDGKAVPQTSPOQHLETFVDGKTYVVDANGORLKNYSWVIDGKTY	180	
Qy	181	FDGQTGEAQTDLPKTGOANQDNVPSQYQANNQAYSNEASSPETVDNYLTADSWSYPRKIL	240	
Db	181	FDGQTGEAQTDLPKTGOANQDNVPSQYQANNQAYSNEASSPETVDNYLTADSWSYPRKIL	240	
Qy	241	KNGSQWQASSEGDLRPIILMTWMPDAATKAAAYANFWAKEGLISGSYRQNSANLDAATQNIQ	300	
Db	241	KNGSQWQASSEGDLRPIILMTWMPDAATKAAAYANFWAKEGLISGSYRQNSANLDAATQNIQ	300	
Qy	301	SAIEKKIASEGTNWLRDKMSQFYKQSQWISIASENETVYPNQDHMOGALLFSNSKQTE	360	
Db	301	SAIEKKIASEGTNWLRDKMSQFYKQSQWISIASENETVYPNQDHMOGALLFSNSKQTE	360	
Qy	361	HANSWRLLNRNPTFQTKQKYFTTNAGYELLANDVDSNPVQAEQLNHLHYLMWG	420	
Db	361	HANSWRLLNRNPTFQTKQKYFTTNAGYELLANDVDSNPVQAEQLNHLHYLMWG	420	
Qy	421	DIVMGDKDANFDGVRVDVAVDNNVADLLQIQRDYKAKYGTQDQEKNAIDHLSILEAWGN	480	
Db	421	DIVMGDKDANFDGVRVDVAVDNNVADLLQIQRDYKAKYGTQDQEKNAIDHLSILEAWGN	480	
Qy	481	DNDYVKDQNNFSLSDNDQORSGLKAFGYSASYRGNLSNLATAGLKNRSANPDSDPVNY	540	
Db	481	DNDYVKDQNNFSLSDNDQORSGLKAFGYSASYRGNLSNLATAGLKNRSANPDSDPVNY	540	
Qy	541	VPIRAHDSVQTRIAKITREKLGKTNADGLTNLTDDLKAPDIYNQDMNATDKVYPNN	600	
Db	541	VPIRAHDSVQTRIAKITREKLGKTNADGLTNLTDDLKAPDIYNQDMNATDKVYPNN	600	
Qy	601	LPMAAYMWLQNKDTVTRYVYGDYTDNQYMATKTPFYNAIETLLKGRIKYVAGQAVSY	660	
Db	601	LPMAAYMWLQNKDTVTRYVYGDYTDNQYMATKTPFYNAIETLLKGRIKYVAGQAVSY	660	
Qy	661	KQDWSSGILLTSVRYGKGANSASDAGNTETRNSGMALLINNRPNFRAYRNLTLNMGAAHKS	720	
Db	661	KQDWSSGILLTSVRYGKGANSASDAGNTETRNSGMALLINNRPNFRAYRNLTLNMGAAHKS	720	
Qy	721	QAYRPLLLSTXKGATYLNDSVDVSRQYKYTDSQGNLSFSASELOSANAQVSGMIQVW	780	
Db	721	QAYRPLLLSTXKGATYLNDSVDVSRQYKYTDSQGNLSFSASELOSANAQVSGMIQVW	780	
Qy	781	PVGAADNQDVTSPTSQATKQGNIXHQSALDSQVIYEGFSNFQAFQSPDQYTNNAVIAK	840	
Db	781	PVGAADNQDVTSPTSQATKQGNIXHQSALDSQVIYEGFSNFQAFQSPDQYTNNAVIAK	840	
Qy	841	NGDLFKSWGITQFEMAPQYVSSDGTFFLDVTLNGYAFSDRYDLAWSKXNKYGSKQDLAN	900	
Db	841	NGDLFKSWGITQFEMAPQYVSSDGTFFLDVTLNGYAFSDRYDLAWSKXNKYGSKQDLAN	900	

use as long as its content is in no way modified and this statement is not removed.

CC EMBL; M29296; AAA26895.1; -; Genomic DNA.
CC EMBL; D88653; BAA26103.1; -; Genomic DNA.
CC EMBL; D88653; BAA26107.1; -; Genomic DNA.
CC EMBL; D88659; BAA26111.1; -; Genomic DNA.
CC EMBL; D88662; BAA26115.1; -; Genomic DNA.
CC EMBL; D89979; BAA26121.1; -; Genomic DNA.
CC EMBL; AB014932; RAN58619.1; -; Genomic DNA.
CC HSSP; P06653; 1GVN.
CC InterPro; IPR002479; Cell wall bd put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 4.
CC Pfam; PF02324; Glyco_hydro_70; 1.
KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
KW Transferase.
FT SIGNAL 1 ? Potential.
FT CHAIN ? 1462 Glucosyltransferase-S.
FT REPEAT 1232 1295 1.
FT REPEAT 1296 1359 2.
FT REPEAT 1360 1423 3.
FT REGION 1232 1423
FT VARIANT 10 10
FT
FT VARIANT 19 19
FT
FT VARIANT 58 58
FT VARIANT 68 68
FT
FT VARIANT 81 81
FT
FT VARIANT 113 113
FT VARIANT 122 122
FT
FT VARIANT 132 132
FT VARIANT 135 135
FT VARIANT 137 137
FT
FT VARIANT 202 202
FT VARIANT 255 255
FT VARIANT 275 275
FT
FT VARIANT 288 288
FT
FT VARIANT 301 301
FT VARIANT 313 313
FT
FT VARIANT 317 317
FT VARIANT 328 328
FT VARIANT 350 350
FT
FT VARIANT 628 633
FT VARIANT 688 688
FT VARIANT 726 732
FT VARIANT 726 730
FT
FT VARIANT 762 762
FT
FT VARIANT 964 964
FT VARIANT 1019 1019
FT
FT VARIANT 1059 1060
FT VARIANT 1060 1060
FT VARIANT 1080 1080
FT VARIANT 1142 1142
FT VARIANT 1198 1198

FT VARIANT 1220 1220 Y -> C (in strain MT4251 and strain
FT MT4467).
FT VARIANT 1280 1280 F -> L (in strain MT4467).
FT VARIANT 1282 1282 Q -> P (in strain MT4245).
FT VARIANT 1290 1290 K -> T (in strain MT4245).
FT VARIANT 1311 1311 N -> D (in strain MT4245).
FT VARIANT 1403 1403 G -> D (in strain GS-5 and strain
FT MT4467).
FT VARIANT 1425 1425 G -> R (in strain GS-5).
FT VARIANT 1449 1449 R -> K (in strain MT4467).
FT CONFLICT 1428 1462 RYDKNSGNMYNKVVTILANGRRIGIDRWGIARY -> VY
FT R (in Ref. 1).
SQ SEQUENCE 1462 AA; 163388 MW; CE4A279C4D708645 CRC64;
Query Match 43.9%; Score 3172; DB 1; Length 1462;
Best Local Similarity 45.8%; Pred. No. 4.2e-14;
Matches 679; Conservative 219; Mismatches 420; Indels 166; Gaps 38;
QY 1 MEKRLRYKLVKHKVKKQWVAIGVTT--VTLSFLAGGVVAADTNNDGTSV-----QVNKMV 53
DB 1 METKRYKQYKVKHWHVTVIAGSLITLGTTLTGSSVSAETEQQTSKDVVTQKSEDDKAA 60
QY 54 PSDPKFDA-----QAQNGLAQAMFKA-----ANQADQATTSQVS 88
DB 61 SESSQTDAPKTKQAQTEQ--TQAQSQANVADTSTSIKTPSQNITQTQNSDDKVTN--T 117
QY 89 PATDGRVNDVTPAANQPAANVANQVANPATDAGALNRQSAADTS-----TGKAVPQTS 145
DB 118 KSEEAQTSSEERTKQAEAAQATASSQALTOAKAEL--TKORQTAAQENKPNVDLAAIP-- 172
QY 146 QPGHLETVDGKTYVYDANGORLKNYSWIDGKTYVFDGQTGCAQTDLN-----KTG 201
DB 173 ---NVKQIDGKYYIYGSGQPKNFALVNNKLVYFDKNTG--ALTDTSQYQPKQGLTKLN 228
QY 202 NVPSQYQANNOAYSNEASPFETVDNYLTADSWYRPRKILKNGQSQHQAASSEGDLRILMTW 261
DB 229 N---DYTHNQIVNPFNTSLETIDNYVTADSWYRPKDILKNGKWTASSEDLRPLLSMW 285
QY 262 WPDAAATKAAAYANFWAKEGLISG--SYRONSA--NLDAATQNTQSAIEKKIASEGNTNWLR 318
DB 286 WPDKQQTAYLNYNNQQGLGTGENYTDSSQESLMLAAQTQVKIETKISQOTQOTQWLRD 345
QY 319 KMSQPVKSONQWSTASENETVYPQDQMHQGGALLPFSNKSQDTEHANSKDRLLNRNPTFG 378
DB 346 IINSFVKTQPNWNSQTESDTSAGEKDKHQGGALLYSNKSXTAYANSDFRLLNRTPTSQG 405
QY 379 KQKVFNTNYA--GYELLANDVNSNPVYQABQNLHLHLMNWGDIVMGDKQANFDGVRVD 437
DB 406 KPKYFEDNSSGGYDFLLANDIDNSNPVYQABQNLHLMNWGDIVMGDKQANFDGVRVD 465
QY 438 AVDNVNADLLOIQRDYYKAKYGTQDNEKNAIDHLSILEAWGNDNDYVKDQNNFSLSDIN 497
DB 466 AVDNVNADLLOIQTASDYKHAHYGVDKSEKNAINHLSILEAWSDNDPQYNNKDTKGAOLPIDN 525
QY 498 DQSCMLKAF-----GYASAVRGNLSNLATAGLKNSA--NPDSDPVNPVVFTRAHDS 548
DB 526 KLRUSLLYALTRPLEKDSNKNNEIRSGLEPVITNSLNNRSABGKNSRMANVIFTRAHDS 585
QY 549 EVQTRIAKIIREKLGKTNADGLTNLTDLNKAFDIYNQDMNATDKVYYPNNLPMAYAMW 608
DB 586 EVQTVIAKIIKQAI--NPKTDLGL--FTLDELKQAFKIYNEDMRQAKKKTQSNIPYAYALM 643
QY 609 LONKDTVTRVYVYGDYTDNGQPMATKTPFFYNAIETLLKGRIKYVAGGO--AVSYKQ---- 662
DB 644 LSNKDSITRLVYGDYSDGQYMATKSPYDAIDTLLKARIKAYAGGQDMKITTYVEGDKS 703
QY 663 ---DWS--SGILTSVRYKGNASASDAGNTETRNSGALLINRPNFRVNRN--LTNLMGAA 717
DB 704 HMDWDYTGVLTSVRYGTGANERATDQSEATKTQGWNAITSSNNPSLUKQNDKVIYVMGTA 763
QY 718 HKSQAYRPLLSTKDGATYLNDSVDVSRQYKTYLSTQGNLSFSASELOSANVAQVSGMITQ 777
DB 764 HKQNEYRPLLSTTKDGLTSYTSDAAKS--LYRKTNDKGELVFDASDIQGYLNPQVSGYLA 822


```

Db      840 KXTDANGILTFDMNDIAGSYNQVSGYLAVWVPVGAKEQNDARVATSKKKKNASGQVYESS 899
QY      809 DALDSQVIYEGSNFOAFQSPDQYTNNAIKNAGDLFKSWGITQFEMAPQYVSSSDGTPL 868
Db      900 PALDSQLIYEGSNFODFATRDQYTNKAIKNVNLFKEWGVTSPQLPQYVSSQDGTPL 959
QY      869 DSVILNGYAFSDRYDLAMSKNNKYSKQDLANAIKGLQSAGIKVLSDLVNPOLYNLPCKE 928
Db      960 DSIQNGYAFEDRYDAMSKNNKYSGLDLDLALRALHSVNTQALADWPDQIYNLPCKE 1019
QY      929 VVTATRVNGOYQAKSGATITNKTPYVANTRSYG-DYQEOYGGKFLDLDLQKLYPRLFTSKQI 987
Db      1020 VVTATRVNNGYTYREGABIEKELYVANSKTNGTDYQKYGGAFDLDELKAKYBEIPEVQI 1079
QY      988 STGKPIDPVSKTINNSAKYFNNGSNILGRGAKYVLSB-GNK-YNLADGKFLFTVLNNTY 1045
Db      1080 SNGQKMTTDEKITKWSAKYFNNGTILGRGAYVYLKDWGSKYLSNKGSTALPKQLVN-- 1137
QY      1046 GQPOVSANGFISKNGGIHYLDKNGQEVKNRP-KEISGSWYVFDSDGKMATGKTIGNDTY 1104
Db      1138 ---KEASTGFVKDTNGFKFYSTSGNOAKDTFIQDENGWNYVFDNOGYLVTGAREIDGKOL 1194
QY      1105 LFWPNGKQLKEGVWYD-GKKAYYDDNGRTWTKNGFVEPRVDGQDKWRYFNGDGTIAIGL 1163
Db      1195 YFMKNGVQLRDALQEDENGQYYYDKTGAKVLNRYTS---DQON-WRYFDAGVMWARGL 1250
QY      1164 VSLDNRTLFDAYGYQVKGQTVTI-NGKSYTFDADGDLVQTDNANPAPQOAGWKLLGD 1222
Db      1251 VKIGDQQYFDQNGYQVKGKVRKADGKLYRFDKDSGNVAVINRFAQ-----GD 1298
QY      1223 N--QWGY-RKDGQLLTGEQITDGQKVFQDNGGVQVKGKGTATDASGVLYRFDYRDOGHQVCK 1279
Db      1299 NPSDMYTFGADGVALTGLQIKIQOQTLFYGDQKQVKGQVVMVLADSKSIRYFDANSGBMAVN 1358
QY      1280 GWYSTDNNWYNVNESGQVLTGLQITDGQTVYFDDKGIQAKGKAVMDENGLRYFDADSG 1339
Db      1359 KFAEGAKNEWYFDQDGKAVTGKLTINNVLVFDQDGKQVGVTLADSKSIRYFDANS 1418
QY      1340 NMLDRW-KVNDGNWYFNRNGLA 1362
Db      1419 EMVANKFAEGAKNEWYFDQDGKA 1442

RESULT 5
Q9LCH3 STOR
ID Q9LCH3 STOR PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfr;
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OC Streptococcus
OX NCBI_TaxID=1303;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC10557;
RX MEDLINE=202311779; PubMed=10768934;
RX DOI=10.1128/IAI.68.5.2475-2483.2000;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RT encoding glucosyltransferase from Streptococcus oralis."
RL Infect. Immun. 68:2475-2483(2000).
DR EMBL; AB025228; BAA95201.1; -; Genomic_DNA.
DR HSSP; P06653; 1HCX.
DR GO; GO:0016740; F:transferase activity; IEA.
DR DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco hydro.
DR Pfam; PF01473; CW binding 1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.

```

```

KW Transferase.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26BAD7C2E543 CRC64;

Query Match 42.7%; Score 3088; DB 2; Length 1575;
Best Local Similarity 44.0%; Pred. No. 6.3e-143;
Matches 669; Conservative 220; Mismatches 456; Indels 176; Gaps 33;

QY 1 MEKNLRYKLHKVKQWVAIGVTTVLSF-----LAGQVVAADTN----- 40
Db 2 MEKIHVKHVKQWVAIALTTLIVAPKVLGLESQVHVHADDVKQVVVQBPATAQTS 61
QY 41 -----NNDGTSSVQNKVWVSPDKPD 60
Db 62 PQQTPPAQKIASQEAKEKVPADKVTDDVAASEKPAKPAENTEATVQNAQEPKPADT 121
QY 61 AQOQNGOLAQA-MFKAANQ-----ADQATTSQVSPATDGRVD-NQVTPAANQAPAA 109
Db 122 KEASTEKAABAEVKAANAITEIPKTEVADQNKQARPTTAQOQEGDKREKTADEVKIVAN 181
QY 110 --VANQD-VANPATDAGALNRQSAADTSTDGKAVPQTSQDOP-----GHLETVDGKTYVDA 162
Db 192 PKVAKKDLPEPGSKQGAIAERMVAD---QAQAPVADHDDDLVLSHIKTIIDGKNYYVD 238
QY 163 NGORLKNYSWIDGTYVFDGQGTGEAQTDLPKTGOANQDNVDPDSYQANNOAQASNEASSPE 222
Db 239 DGTVKCNFAVELNGRILYFDAETGALVDSNEVFOQGTSSLNNEFSQKNAFYGTDDKDI 298
QY 223 TVDNYLTADSWYRPRKILKNGSQWASSEGDLRPLMTWPDAAATKAAAYANFWAKEGLIS 282
Db 299 TVDGYLTADSWYRPRKILKNGSQWASSEGDLRPLMTWPDAAATKAAAYANFWAKEGLIS 358
QY 283 GSY--RQNSANLDAATQNIQSAIEKKIASGNTWLRDKMSQFVKSQNQSASENEITY 340
Db 359 GAFENKVEQALLTGASQVQRKIEKIGKEGDKWLRTLMGAFVKTQPNWNITKTESYTG 418
QY 341 PNQDHMOGALLFNSKDETHANSQWRLNLRNPTFTQTKQKYF-TTNVAGYELLANDVD 399
Db 419 TKDHLQGGALLYTNNEKSPHADSFRLLNRTPTSTQTPKYFIDKSGNGYFELLANDFD 478
QY 400 NSNPVVAEQQLNHLHYLMNWGDI VMGDKDANPDGVRVDAVNVNADLLQIQRIQRIYKAKYG 459
Db 479 NSNPVVAEQQLNHLHYLMNWGDI VMGDKDANPDGVRVDAVNVNADLLQIQRIQRIYKAKYG 538
QY 460 TDQNEKNAIDHLSILEANSGNDNDYVQDNFSLIDNDORSGLMKAFQVAAAYAGNLSN 519
Db 539 VGESEEAIKHLSILEANSGNDNDYVQDNFSLIDNDORSGLMKAFQVAAAYAGNLSN 598
QY 520 LATAGLKNRSA-NPDSDPVNPVYFIRAHDSQVQTRIAKIRKLGKTNADGLTNLTDDL 578
Db 599 TTITNSLDRSEKNGERMANYIFVRAHDSQVQTRIAKIRKLGKTNADGLTNLTDDL 656
QY 579 NKAFDINQDMNATDKVYPPNNLPMAYAMLMQNDQTVTRVYVYDGMVYDNGQYMATKTPPY 638
Db 657 KQAFKIYNEEDMRKADKKYQFNIPTAHALMNSKDSITRVYVYDGLYDQDQWMEKSPYH 716
QY 639 NAIETLLKRIKYVAGGO--AVSY-----KQDWS-SGILTSVRYCKGANSASDAGNTET 689
Db 717 DAIDALLRARIKYVAGGQDMKYTVNGVPREADKWSYNGILTSVRYCTGANEATDEGTAET 776
QY 690 RNSGMALLINNRPNFR--AYRNLTILNMGAHKSQAYRPLLLSTKQGIATYLANDSDVDSRQ 747
Db 777 FTQGMVATSNPNLKLNEWDKLVNMGAAHKNQYRPLLTKDQISYLLDEEVPQSL 836
QY 748 KYKTDSQGNLSFSASELOSVANAQVSGMIQVWVPVGAADNDQVRTSPSQATKQGNIIHQ 807
Db 837 WKKTANGILTFDMNDIAGSYNQVSGYLAVWVPVGAQADQARTTASKKKNASGQVYES 896
QY 808 SDALDSQVIYEGFSNFOAFQSPDQYTNNAIKNAGDLFKSWGITQFEMAPQYVSSSDGT 867
Db 897 SAALDSQVIYEGFSNFOAFQSPDQYTNNAIKNAGDLFKSWGITQFEMAPQYVSSSDGT 956
QY 868 LDSVILNGYAFSDRYDLAMSKNNKYSKQDLANAIKGLQSAGIKVLSDLVNPOLYNLPCK 927

```


Db 1000 DMFTVMNISTGKPIDSTKIKQWEAKYFNGTIVNLGKAGYVLSDDATGKYFTVNGDFL 1059
QY 1038 PTVLNTTYQPOVSANGFTISKNGGIHYLDKNGOEKVRNKEISGSHYYPDSGKMATG-- 1095
Db 1060 PA-----SFTGDQNAKTGFYDGTGMAYYSTGNKAVNSFIYEGGHYYPDKDGHMVTGSY 1115
QY 1096 KTKIGNDTYLPWPNKGQKLEGVYDGK-KAYYYDDNGRWTWKG-----FVEFRVDGQDK 1149
Db 1116 KAEDGND-YFPLNGIQMRDAIYQDAQNSYIYIGRTGILY--KGNWYFFVDPNNANKTV 1172
QY 1150 WRYFNGDGTIAIGLSLNRITLYFAYGYQVKGQVTINGKSYTFDADQG----- 1199
Db 1173 FRYEDANNWAIGYRMYGQTYFFBENGFAQKQLLTDKGTGFDEDNAMAKNFVNV 1232
QY 1200 --DLVQTDNANPAPOGQ----- 1214
Db 1233 GDDWYTMGNGNAVKGQYVWNQILYFNPETGVQVKGQFITDAQRTSYDANSALKSS 1292
QY 1215 -----AGWKLLGDN--ONGY--RKDGQLLTGEQITIDGQKVEFO-DN 1250
Db 1293 GFPTNGSDWYEAENGYYVKGFKQVAENODQWYFDQTTGKQAKAAKVDGRDLVFNPD 1352
QY 1251 GVOVKGTADASGLRFRFYDRQGHQVKGWYSTSDNNWVYVNESQVLTGLQITID---- 1306
Db 1353 GVOVKGDFATDESGTFSFYHGDNGDKVVGFFTTGNNAWYADNNGNLVKGFQEI DGWY 1412
QY 1307 -----GQTVYFD-DKGIQAKGKAVMDENGLRYFDDADSGNMLDRWKN 1348
Db 1413 HFDEVTGQAKGAALVNGQQLYFDVDSGIVQVKGDFVTDQGNSTYYDVNSGD-----KK 1466
QY 1349 VDG-----NWYFNRNG 1360
Db 1467 VNGFTTGDNWYADGQG 1485

RESULT 7

OS5264_STREL PRELIMINARY; PRT; 1449 AA.
AC OS5264;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Name=gfll;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]_TaxID=1304;
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.I., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR EMBL; J35495; AAC41412.1; -; Genomic_DNA.
DR PIR; T30857; T30857.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
FT SIGNAL 1 35 Potential.
FT CHAIN 36 1449 Glucosyltransferase.
FT SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 42.6%; Score 3079.5; DB 2; Length 1449;

Best Local Similarity 45.0%; Pred. No. 1.5e-142;

Matches 667; Conservative 206; Mismatches 451; Indels 157; Gaps 35;

QY 1 MEKNRLYKHLKVKQWVAIGVTTVTLSPLAGQVVAADTNNNDGTSVQVKNKWPSPDKFD 60

Db 1 MDKVVHYKHKVKQWVIAVTGLSL-----GAVSAVSLGTNDG-----VQADEHTD 48
QY 61 AQAQNGLOAAMFKAANO--ADQATATSOVSPATDGRVDNQVTPAA-----NQPAAN 109
Db 49 ATVAIPDITVDTGTVSNDTAAQDPTTAAATNDVATD-QATPTATFDLTTDTTNTVAAN 107
QY 110 VANQDVANPATDAGALNRQSA-----DTSIDGKAV----- 140
Db 108 -AVDTVATVGTDRAAATTNDTATNDTAVDTTNNNTTTTDTTVDRAATERRATGARROP 166
QY 141 -----POT-----SDQGHLETVDGKTYVVDANGORLKN 169
Db 167 TCGRRATPVNGTNNWNTTVVNNDLPATNNVVDGFSHIKTINGKQYVVEDDGTIRKN 226
QY 170 YSMVJDKTYYPDGTGEAQT-----LPKTOANQDNVPSYOAN-----NQAYSNEA 218
Db 227 VYLERIGSGQYFNAETGELSNOKEYRFDKNGGTGSS--ADSTNTNTVTVNGDKNAFVGTTD 284
QY 219 SSFETVDNYLTADSWYRPRKILKNGQSWQASSEGDLRPIILMTWTPDAATKAAAYANFAWE 278
Db 285 KDIELVDGTYFTANTWYRPRKEILKDGKWTASTENDKRPULLTVWVWFSKAIQASLYLNMKEQ 344
QY 279 GL-IGSYRONSAA--NLDAATONIQSAIEKKIASEGNTNWLDRKMSQFVKSONQWSIAS 335
Db 345 GLGTTQVTTSSQTMQDAALEVKRLEERLAREGNTDMLRTTIKNFVKTPQGHNSISE 404
QY 336 NETVYPNODHMQGALLFSNSKOTEHANSWRLLNRNPTFQTKQ--KYF-TTNYAGYEL 392
Db 405 N---LDNNDLHQQGALLYNNDSTSHANSRYLLNRTPTSTQTKENPKYTKDTSNGPPEF 461
QY 393 LLANDVDSNPVVQAEQLNHLHYLMNWGDI VNGDKDANFDGVRVDAVDNWNADLLQIQRD 452
Db 462 LLANDIDNSPAVQAEQLNWLHYIMNIGITGSGEDENFDGVRVDAVDNWNADLLQIASD 521
QY 453 YYKAKYGTDONEKNAIDHLSTLEAWSGNDNDVVKDQNNFSLSDNDQSGMLKAFGYASA 512
Db 522 YFKAKYGADQSQDAIKHLSTLEAWSHNDAYYNETDKAQLPMDPMDHMLALYSLLRPTG 581
QY 513 YRGNLSNLATAGLKNRS--ANPDSDPVNVYFIRAHDSVQTRIAKIIRKLGKTNADGHT 571
Db 582 NRSQVEPLISLNDRSSESGKSKRMANYAFVRAHDSVQSIIQIIKNEINPQSTG--N 639
QY 572 NLTLDDLAKAFDIYNQDNATDKVYYPNNLPMAYAWMLONKDTVTRVYVYGDYTDNGQYM 631
Db 640 TFTLDEMKAFEIYNKQWRSANKQYTNIPISAYALMLTKDTPVPRVYVYGDYTDGQYM 699
QY 632 ATKTPFYNAIETLLKGRIKYVAGGO--AVSY-----KQDW-SSGILTSVRYGKANSASD 683
Db 700 AOKSPYDAIETLLKGRIRYAAAGGQDMKVNYIGYNTNGWDAGVLTSVRYGTGANSASD 759
QY 684 AGNTRNSGMALLINRPNFRAYRNLTLNMGAAHKSQAYRPLLSTKDGIAITYLNDSDV 743
Db 760 TGASTRNQGMAVIVSNQPALRLTSLNLTINMGAAHRNQAYRPLLLTNDGVATYLNDSDA 819
QY 744 DSRQKYTDSQGNLSFSASELSQVANAQVSGMIQVWVPVGAADNODVRTSPSTQATKDN 803
Db 820 NG-IVKYTDGNGNLTFSSANEIRGINPQVDGYLAWVPVPGASENQDVRVAPSKENSSGL 878
QY 804 IYHQSDALDSOVIYEGFSNFQAFQSPDQYTNNAVIAKNGDLFPKSWGITQFENAPQVVSSE 863
Db 879 VYESNAALDSQVIYEGFSNFQFVQNPQSYTNKKIAENANLEPKSWGITSFEPAPQVSSD 938
QY 864 DGTFLDSVLNGYAFSDRYDLAMSNNKYGSKQDLANAIKGLQSAIGIKVLSDLVNPQLYN 923
Db 939 DGSFLDSVIQGYAFTDRYDIGMSKDKYGLADLKAALKSLHVAVGISAIAADWPQIYN 998
QY 924 LPGEKVVTTATRVNQYQAKSGATINKTPYVANTRSYG-DYQEQYGGKFLDLDLQKLYPRLF 982
Db 999 LPGEVWVTTATRVNNYGETKQGAIDHSLYAAKTRTFGNDYQGYGAFDELKRLYPPQIF 1058
QY 983 STKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLSG-NKYLNLADGKLFJPTVL 1041

Db 1059 DRVOISTGKRMWTTDEKIQWSAKYMKMGNTNILDNRGSEVYLKNGLNGYVGTNGGKVSPLPKVV 1118
Qy 1042 NNTYGOQVVSANG-----FISKNGIHYLDKNGQEVNRF-KEISGSWYFFSDGKMATG 1095
Db 1119 GSNQSTGNDGNGDGGSKFEKRLFSVRYRYNNGQYAKNAFIKNDGNGVYFFDNGRMAVG 1178
Qy 1096 KTKIGNDTYLFMPNGKQKLGWYDVGK-KAYYYDDNGRTWTN-----KGFVE 1141
Db 1179 EKTIDGKQYFFLANGVQLRDGYRNRGQVFFYDQNGVLNANGKQDKPKDNNNNSGRNQ 1238
Qy 1142 FRVDGDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGVQVKGQVTTNGKSYTFDADQGLD 1201
Db 1239 FVQIGNWYAYDNGKRVTHQNINGQELFFDNNNGVQVKGRTVNGAIRYDANSSEM 1298
Qy 1202 VQTDNANPAPQGOAGMKLLGDNQWGY-RKDQQLLTGEQTTIDGQKVFQDNGVQVKGGTAT 1260
Db 1299 ARNRFAEIEP-----GWAYFNNDGTAVKGSQNINGQDLYFDQNGRQVKGALA- 1346
Qy 1261 DASGLVLFYDRDQHQVGKQWYSTSDNNWYVNESGOVLTLGLOTIDGQTVYFDKGIQAK 1320
Db 1347 NVDGNLYRYDVNSG-ELYRNRPEHIDGSWTYFDGNGNAVKGMYNINQONLLFDNNGKQIK 1405
Qy 1321 GKAVWDENGLRYFDADSGNMLDRWKNVD-GNMYVFNRRNG 1360
Db 1406 GHUV-RVNGVVRYPDNGEMAVNRWVEVSPGWWVYFDGEG 1445

RESULT 8

O68542_STRSL PRELIMINARY; PRT; 1449 AA.
ID O68542; AC O68542; DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase N (Fragment).
GN Name=gtfN;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
RN NCBI_TaxID=1304;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=V1477;
RA Jaffe R.I.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049609; AAC05156.1; -; Genomic_DNA.
DR PIR; T30552; T30552.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 2_
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
FT NON_TER 1449 1449
SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;

Query Match 42.6%; Score 3076.5; DB 2; Length 1449;
Best Local Similarity 45.0%; Pred. No. 2.1e-142;
Matches 667; Conservative 205; Mismatches 452; Indels 157; Gaps 35;
Qy 1 MBKNRYKLHKVKKQWVAIGVTVTLSFLAGGQVVAADTNNDGTSVQVKNWVPSPDKFD 60
Db 1 MDKKVHYKHVKQWVAIGVTVTLSFLAGGQVVAADTNNDGTSVQVKNWVPSPDKFD 60
Qy 61 AQANQQLAAQMFKAANQ--ADQTATSQVSPATDGRVDNQVTPAA-----NOPAAN 109
Db 49 ATVAIPDITVDGTGVSNDTTAAQDPDTAAATNDVATD-QAIPATFDLTDTTNTVAAN 107
Qy 110 VANQDVANPATDAGALNRSA-----DTSTDGKAV----- 140
Db 108 -AVDTVAVTGTDRAATTNDTTATNDTAVDTTNNNTTTDTTNTNRAATTERRATGARRGP 166

Qy 141 -----POT-----SDQPGHLETVDGKTYVVDANGORLKN 169
Db 167 TGGRRATPVNGTNNANNTVVVNNDLPATNNVVTDGPHIKTINGKQYVVEDDGTIRKN 226
Qy 170 YSNVIDGKTYYPGQGTGEAQT-----LPTGQANQDNVPDSYOAN-----NQAYSNEA 218
Db 227 YVLERIGGSOYFNAETGELSNOKEYRFDKNGGTGSS--ADSTNTNTVNGDKNAFYGTITD 284
Qy 219 SSPETVDNTLTADSWRPRKILKNGOSWQASSBGDLRPIILMTWMPDAATKAAAYANFWAKE 278
Db 285 KDIELVDGFTANTWRPKELLKDGKEWTASTENDKRPLLTVMWPSKAIQASLYNTMKEQ 344
Qy 279 GL-IGSRYRNSA--NLDAAATQIQSAIEKKIASEGNTWLRDKMSQFVKSNQWMSIAS 335
Db 345 GLGTNTYTSFSSQTMQDALEVKRIEGRIAREGNTDWRRTTIKNFVKTPQGMNSTSE 404
Qy 336 NETVYENQDHMOGGALLPSNSKQTEHANSQDWLRLLNRNPTFQCKQ--KYP-TTNYAGYEL 392
Db 405 N-----LDNNDHLOGGALLYNNDSRTSHANSDYRLNRTPTTSQTKHNPKYTKDTSNGGF 461
Qy 393 LLANDVDNSNPVQABQLNHLHYLMNWGDI VMGDKDANFDGVRVDAVDNVDNADLLQIORD 452
Db 462 LLANDIDNSNPVQABQLNHLHYIMNIGITGSEDEDFGVRVDAVDNVDNADLLQIASD 521
Qy 453 YYKAKYGTQDNEKNAIDHLSIILEAMSGNDNDYVQDNFSLSIDNDQSRGMLKAFGYASA 512
Db 522 YPKAKYGADQSDQAIKHLIILEAWSHNDAYYNEDTKGAQLPMDPMHLALVYSLLRPIG 581
Qy 513 YRGNLSLATAGLKNRS--ANPDSDPVPNTVPTTRAHDSSEVQTRIAKILREKLKTNADGLT 571
Db 582 NRSQVEPLTNSLNDRSSESGKSKRMANTAFVRAHDSSEVQSIIGQIKNEINPQSTG--N 639
Qy 572 NLTLDDLKNAFDIYNQDMNATQVYVPPNLPWAYAMLONKDTVRYVYGDWYTDNGOYM 631
Db 640 TTLDLMDKKAFFLYNKMRSANKQYQYINPISAYALMLTHKQTPRVYVYGDWYTDGQYM 699
Qy 632 ATKTPFYNAIETLLKRIKYVAGGQ--AVSY-----KODW-SSGILTSVRYKGSANSASD 683
Db 700 AQKSPYDAIETLLKGRIRYAGGQDMKYNIGYGTNGWDAGVLTSVRYGTGANSASD 759
Qy 684 AGNETRNSGSMALLNNRPNFRAYRNLTLNMGAAHKSQAYRPLLLSTKQGIATYLNDSOV 743
Db 760 TGTAETRNQGMVIVSNQPALRLTSLNTINMGAAHNRQAYRPLLLTNDGVATYLNDSDA 819
Qy 744 DSRQVKYTDSCNLSFSAELOSANAQVSGMIVQVWVPVGAADNQDVRTSPSQATKDN 803
Db 820 NG-IVKTYDNGNLTFSANEIRNPQVDGYLAVWVPVGAASENQDVRVAPSKERNSSGL 878
Qy 804 IYHQSDALDSQVIYEGFSNFQAFQSPQOYTNVNAIKNGDLFKSMGITOFEAPQVVSSE 863
Db 879 VYESNAALDSQVIYEGFSNFQDFVNPSPQYTNKKIAENANLFSKWGITSPFAPQIVSSD 938
Qy 864 DGTFLDSVLNGYAFSDRYDLAMSKNKKYSGKODLANAIKGLQSAKIKVLSLDPNQLYN 923
Db 939 DGSFLDSVTQNGYAFTRDYDIQMSKDNKYGLSLADLKAALKSLHAVGISAIADWVPDQIYN 998
Qy 924 LPGEKVVATRVNOYQOAKSGATINKTPVAVNTRSYG-DYQSYGSKFLDDQLKLYPLRF 982
Db 999 LPGEDEVVATRVNNTGETKDGAIIDHSLYAAKTRTFGNDYQGYKGFADLRLKLYPQIF 1058
Qy 983 STKQISTGKPIDPSVKITNWSAKYPNGSNILRGARYVLSEK-NKYLNLNADGKLFPTVL 1041
Db 1059 DRVOISTGKRMWTTDEKIKWSAKYMKMGNTNILDNRGSEVYLKNGLNGYVGTNGGKVSPLPKVV 1118
Qy 1042 NNTYGOQVVSANG-----FISKNGIHYLDKNGQEVNRF-KEISGSWYFFSDGKMATG 1095
Db 1119 GSNQSTGNDGNGDGGSKFEKRLFSVRYRYNNGQYAKNAFIKNDGNGVYFFDNGRMAVG 1178
Qy 1096 KTKIGNDTYLFMPNGKQKLGWYDVGK-KAYYYDDNGRTWTN-----KGFVE 1141
Db 1179 EKTIDGKQYFFLANGVQLRDGYRNRGQVFFYDQNGVLNANGKQDKPKDNNNNTSGRNG 1238
Qy 1142 FRVDGDKWRYFNGDGTIAIGLVSLDNRTLYFDAYGVQVKGQVTTNGKSYTFDADQGLD 1201

Db 879 STDQGFELDSVIQNGYAFTDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVDPQM 938
QY 922 YNLPGKEVTVATRNVOYQAKSGATINKTPYVANTRSYG-DYOEQYGGKFLDLQKLYPR 980
Db 939 YTFPKQEVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGGDYQAKYGGAFDLDELKEKYPE 998
QY 981 LFTSKQISTGKPTDPSVKITNWSAKYPNGSNILGRGAKYVLSB--GNKYNLADGKFLFLP 1038
Db 999 LFTKQISTGQAIDPSVKIKOWSAKYPNGSNILGRGADYVLSDAQSNKYNLNSDDKFLFLP 1058
QY 1039 TVL----- 1041
Db 1059 KILLQGVESGIRFDGTGYVNSSTGCKYDTSFITEAGNLVYFGQGVNVTGAQNIKGS 1118
QY 1042 -----NNTYG-----QPQVSANGFISK 1058
Db 1119 NYIFLANGAALRNVTYVTDAGQGNHYTGNDGKRYENGQYQFGNDSWRYFKNGVVALGLTIV 1178
QY 1059 NGGIHYLDK-----NGQEVKNRP-KEISGSWYTFDSGKMAT 1094
Db 1179 DGHVQVFDKGVQAKDKIIVTRDGKRVYFDQHNNAVNTTFVADKTHWYTLGKGVAVT 1238
QY 1095 GKTKIGNDTYLEMPNGKOLKEGVYDCK--KAYYD-DNGRTWTKGFVEFRVGDQKWR 1151
Db 1239 GAQTVGKQHLVFEANGQOVK-GDFTAKOGKLYFDVDSGDMWTNT-FIE---DKAGNWF 1293
QY 1152 YFNGDGTIAIGVLSLNRILTYDAYGVQVKG-----QVTIN 1188
Db 1294 YLKGGAATVGAQTIKGKLYFKANGQOVKGVLDKADGKIRYYDAQTGEQVFNKSVSN 1353
QY 1189 GKSYTFDADGDLVQTDNANPAPQ-----GQ-----AGWLLGDNONGYRK 1229
Db 1354 GKTYFGSD--GTAQT-QANPKGQTFKDGSGVLFYFNLEQYVSGSGWYETAHEHWYVK 1410
QY 1230 DGLLNGEOTIDQKVFQDNGVQVKG----- 1256
Db 1411 SGKVLTAQOTIGNQRYFKDNGHQVQGLVTGNDGKLYRYDANSQDAFNKSVTVNGKTY 1470
QY 1257 -----GTA-----TDASGVLFYDRDQGHQVKGWYSTSDNWNVYNESQVLT 1300
Db 1471 YFGSDTAQTAQNPQKQTFKDGSGVLFYFNLEQYVSGSGWYKNAQOQLYVKD-GKVL 1529
QY 1301 GLQTDGQTVYFDKGIQAKGKAVDENGKLYRYFDADSGMULDRKVNKVDGNYVFNNG 1360
Db 1530 GLQTVGNQKVFYFDKGIQAKGKAVRTSDGKRVYFDENSGSMITNQWKFYGVYVYFGSDG 1589
QY 1361 LA 1362
Db 1590 AA 1591

RESULT 10
Q5983 9STRE PRELIMINARY; PRT; 1590 AA.
AC Q5983;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtf1;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]_TaxID=1310;
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=OM2176;
RC MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2]

RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefna S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; LHXC.
DR GO; GO:0047849; F.dextranucrase activity; IEA.
DR GO; GO:0016757; F.transferase activity, transferring glucosyl. . .; IEA.
DR GO; GO:0009250; P.glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 2_.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 41.9%; Score 3031; DB 2; Length 1590;
Best Local Similarity 41.6%; Pred. No. 4.1e-140;
Matches 672; Conservative 217; Mismatches 436; Indels 292; Gaps 36;

QY 1 MEKNRLKHLKVKKQWVAIGVTVT-LSFLAGGVVAAADTNNDGTSVQVKNKVPDPKPF 59
Db 1 MEKNVRFKHKVKRWTLVSASATWLASALGASVASADTDTASDDSNQ--VVTGD--- 55
QY 60 DAQAQNGQLAQMFAKNAQADQATTSQVSPATDGRVDNQVTPAANQVPAANQV--- 115
Db 56 --QTTNNQATDQTSIAA-----TATSEQASATDAATDQ--ASAAEQTQGTASTDTAAQT 106
QY 116 ---ANPA-----TDAGALNRQSAADTSTDGKAVPQTSDDPGHLETVGDKTYV 160
Db 107 TTNANEAKWVPTENENQGFDEMLAEAKNVATAES--SIPDLAKSNVKKVQDGYKYV 164
QY 161 DANGQRLKNYSVIDKTYIFDQGTG---EAQTDLPKQANQVNDVPDSYOANNQAYSN 216
Db 165 DQGNVKNFVAVSGDKIYFP-ETGAYKDTSEKVDADKSSAVSQVA-TIPANNRAYST 222
QY 217 EASSFETVDNYLTADSWYRPRKILKNQSQWASSBGDLRPLMTWMPDPAATKAAAYANFWA 276
Db 223 SAKNEAVDNYLTADSWYRPRKILKXGKTWTSKGGDDFPELLMAWMPDTEKRYVNYVN 282
QY 277 KEGLISGSY--RQNSANLDAATQNTQSAIEKKIASBGNVNLDRKMSQVFKSQNQWSIAS 334
Db 283 KYVGIDKTYTAETSOADLTAAAEVQARIEQKITSENTKWLREAIASFVKTPQWNNGES 342
QY 335 ENETVYVQNDHMQGALLFNSKD--TEHANSWRLLNRNPTFTQTK-QKYFTTN----YA 388
Db 343 EK----PYDDHLQNGALLFDNQTDLTPDTQSNRYLLNRNPTFTQTKSLDSRFTYNDPLG 398
QY 389 GYELLANDVNSNPVQAEQLNHLHLYLMNWGDIYVMGDKDANFDGVRVDVADVNADLLQ 448
Db 399 GYDFELLANDVNSNPVQAEQLNHLHLYLMNWGDIYVMGDKDANFDGVRVDVADVNADLLQ 458
QY 449 IORDYKAKYGTQDQNEKNAIDHLSILEAWSGNDNDYVDQNNFSLSIDNDQSGMLKAFG 508
Db 459 ISSDYLKAAYGIDKNNKNNHVSIVEAWSNDTPLYLHDDGDNLMNMDKFRLLMSLWSLA 518
QY 509 YASAVRGNLSNLATAGLNRSANP--DSDPVNPVVFIRAHDSVQVTRIAKIIIEKLGKTNW 567
Db 519 KPLDKRSGNLNLIHNSLVDRREVDTVPSTSFARAHDSVQDIIRDIIRKAEI-NPNS 577
QY 568 DGLTNLTLDLKNKAFDIYNQDMNATDKVYPPNPLPMAYAMWLQNKDTVTRVYVYGDYNT 627
Db 578 FGY-SFTQEEIQAQFKIYNEDLKKTDKKYTHYVPLSYLLTLLTKNGSIIRVYVYGDYNT 636
QY 628 GOYMATKTPFYNAIETLLKGRIKYVAGGQAVSYKQDMSSGILTSVRYGKANSASDAGNT 687
Db 637 GOYMAKTVNYDAIESLLKARMKYVSGQAMQNYQITNGEILTSVRYGKALKQSDKGA 696

QY	688	ETNSGSMALLINRRPNERA-YENLTINMGAAHKSQAYRPLLLSTKDGIAATYLNDSVD-S	745
Db	697	TTRTSGVGVVMGNQNFSLDGVKVALNMGAAHANOEYRALMVSTKDGVAITYATDADASKA	756
QY	746	ROYKYTDSOGNLSFSAELQSVANQVSGMIQVWPVGAADNQDVRTSPSTQATKDGNITY	805
Db	757	GLVKRTDENGILYFLNDDLKGAVNPQVSGFLQVWPVGAADQDDIRVAASDTASTDGKSL	816
QY	806	HQSDALDSOVIYEGFSNFOAQSPOQYNAVIAKNGDLFKSWGITQFEMAPQYVSSBDG	865
Db	817	HQDAAMDSRVMEGFSNFOAQSPOQYNAVIAKNGDLFKSWGITQFEMAPQYVSSBDG	876
QY	866	TELDVILNGYAFSDRYDLAMSKNNKYSKQDLANAIGLQSAQIKVLSLVPNQLYNLP	925
Db	877	QFLDSVIQNGYAFSDRYDLAMSKNNKYSKQDLANAIGLQSAQIKVLSLVPNQLYNLP	936
QY	926	GKEVVTATRVNOYQAKSGATINKTPYVANTRSYG-DYQEOYGGKFLDLOKLQYLPRLPST	984
Db	937	KQEVVTATRVNOYQAKSGATINKTPYVANTRSYG-DYQEOYGGKFLDLOKLQYLPRLPST	996
QY	985	KOISTGKPIPSVKITNNKSAKFNKNGNLSILGRGAKYVLS--GNKYNLNLADGKLFPTVLN	1042
Db	997	KOISTGKPIPSVKITNNKSAKFNKNGNLSILGRGAKYVLS--GNKYNLNLADGKLFPTVLN	1056
QY	1043	NTYQPOVSANGFISKNGGHIHYLDKNGQVKNRFXEISGSWYVFDSDGKMATKTKIGND	1102
Db	1057	GKVSSEGRIDG----KGVIYNSSATGDQVQKASFITEAGNLVYFGKDGVMVYGAOTINGA	1112
QY	1103	TYLFPNGKOLKEGVYDGG-KAYYYDNGRTWTKNGFVEFRVDGQKWRVYFNGDGTIAI	1161
Db	1113	NYFFLENGTALRNTIYTDAGNSHYVANDGKEYEN-GYQOF---GND-WRYFK-DGNMAY	1166
QY	1162	GLVSLD-----	1167
Db	1167	GLTIVDGNVOYFDKGVQAKDKIIIVTRDGVRYFDQHNHNAATNTFIADKTHWYVLGKD	1226
QY	1168	-----NRTLYFDAYGVQVKGOTVTIN-GKSYTFDADQGL-----	1201
Db	1227	GVAVTGAOTVKGOKLYFEANGQVVRGDFVTSDEGLYFYDVDSGDMWTDFTIEDKAGNWF	1286
QY	1202	-----	1201
Db	1287	YLGKDGAAVTGAOTIRGQKLYPKANGQVKGDI VKGTDGKIRYVDAKSGEQVFNKTVKAA	1346
QY	1202	-----VOTDANPAP-----	1231
Db	1347	DGKTYVIGNDGVAVDPVS VVKGTQFKDASGALRFPYNLKGQLVTGSGWYETANHDWVYIQSG	1406
QY	1232	QLLTGEOTIDGQVFPDNGOVQVKGATDASGLVRFYDRDQHQ-----	1276
Db	1407	KALTGEOTINGOHLYPKEDGHQVKGOLVGTGDKVRYVDANSQDQAFNKSVTNKGTYYP	1466
QY	1277	-----VKGWYSTSDNNWYVNESGQVLTGLQ	1303
Db	1467	GNDGTAQTAGNPKGTQFKDGSIRFPYSMEGQLVTGSGWYENAGQWLYV-KNGKVLTLGLQ	1525
QY	1304	TIDGQVYFPDDKGIQAKGKAVVDENGLNLYFDADSGNMLRDRWKNVDGNWYFNRRNG	1360
Db	1526	TVGSQRVYFDNGIQAQKAVRSTSDGKIRYFDENSGSMITNQWKFYGYGYFYFGNDG	1582
RESULT 11			
ID	Q55265	STRSL	
AC	Q55265	PRELIMINARY;	PRT; 1577 AA.
DT	01-NOV-1996	(trEMBLrel. 01, Created)	
DT	01-NOV-1996	(trEMBLrel. 01, Last sequence update)	
DT	01-JUN-2003	(trEMBLrel. 24, Last annotation update)	
DE	Glucosyltransferase precursor.		
GN	Name=gtfM;		
OS	Streptococcus salivarius.		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		

OC	Streptococcus.		
OX	NCBI_TaxID=1304;		
RN	(1)		
RP	NUCLEOTIDE SEQUENCE.		
RX	MEDLINE=9512197; PubMed=7822030;		
RA	Simpson C.L., Giffard P.M., Jacques N.A.;		
RT	"Streptococcus salivarius ATCC 25975 possesses at least two genes		
RL	coding for primer-independent glucosyltransferases.";		
DR	Infect. Immun. 63:609-621(1995).		
DR	EMBL; U35288; AAC1413.1; -; Genomic_DNA.		
DR	PIR; T30858; T30858.		
DR	HSSP; P06653; 1H8G.		
DR	GO; GO:0016740; F:transferase activity; IEA.		
DR	GO; GO:0009250; P:glucan biosynthesis; IEA.		
DR	InterPro; IPR004829; Caurface_antigen.		
DR	InterPro; IPR002479; CW_binding.		
DR	InterPro; IPR003318; Glyco_hydro_70.		
DR	Pfam; PF01473; CW_binding_1; 4.		
DR	Pfam; PF02324; Glyco_hydro_70; 1.		
DR	ProDom; PD153432; Caurface_antigen; 1.		
KW	Signal; Transferase.		
FT	SIGNAL	1 38	Potential.
FT	CHAIN	39 1577	glucosyltransferase.
SQ	SEQUENCE	1577 AA; 175290 MM; 3EFB898A7D3A78F3	CRC64;
Query Match			
Best Local Similarity 41.9%; Score 3026.5; DB 2; Length 1577;			
Matches 669; Conservative 215; Mismatches 440; Indels 245; Gaps 42;			
QY	1	MEKNLRYKXKHKQWVAIGVTTVTLSPFAGGQVVA-----ADTN-----NNDG-----	44
Db	1	MENKVRPKLHKVKNWVITIGVTTLSWALAGSLLAQGVKVEADETSAPNGDGLQLSBDG	60
QY	45	-----TSVQVKNMVPSPKFDQAQNGQLAQMFKAAQNA-----DOTATSVQSPATDGRV-	95
Db	61	TASLVTTTVTTEQASQAQSAVSASVASHETSFOATSAVSQEATAQTSFVASQEVA	120
QY	96	-----DNQVTP--AANQPAANVANO-----DVANP--ATDAGALNRQSA	131
Db	121	VSSQTSSGQETQTTEQVSGQTSQVAGQTSAQSTPSTVTEQARPRVLTNAAPATRAA	180
QY	132	D-----TSTDG-----	137
Db	181	DSTIRINARNRNTITITASGITPNTVITIGPNTKPNVTITSPNGTRPNVTIVTQPNQN	240
QY	138	KAV-PQTSQDP-----GHLETVDGKYYVDANGQRLAKNYSMWIDGKTY	179
Db	241	KPVQPSQSPQNKPVQPNQPSLDYKPVASNLKTIIDGKQYYVE--NGVVKNAAIELDGRLY	299
QY	180	YFDGOTGEAQTDLPKTGQANODNVPS--YQANNOAYSNEASSFETVDNYLTADSWYRPR	237
Db	300	YFD-ETG-AMVDQSKP-LYRADAI PNNSIYAVYNOAYDTSSKSFELHDFLTADSWYRPK	356
QY	238	KILKNGQSQWASSEGLRPIILMTWMPDAATKAAYANFWAKEGLISGSYRQN--SANLDA	295
Db	357	QILKDGKWTASTEKDYRPLMTWMPDKVTQVNYLNYMSQQGFGNKTYTTDMMSYDLAAA	416
QY	236	TONTQSAIEKKITASEGNTNWLDRKMSQFVKSONQWSIASENETVYPNQDHOGGALLFSN	355
Db	417	AETVQRGTEERTGREGNTTLRLQMSDFIKTPQPGWNSESE--DNLLVGDHDLQGGALTFLN	475
QY	356	SKDTEHANSRDLNLRNPTFOTGKQY-FTTNYAGVELLLANDVNSPNVQAEQLHLH	414
Db	476	NSATSHANSDFRLMNRTPNTQGTGRKHIDHSNGYFELLANDINDSNPQAEQLNWLH	535
QY	415	YLMNWGDI VMGDKDANFQGVRVDAVDNVNADLLQIQRDYIKAKYGTQDQKNAIDHLSIL	474
Db	536	YIMNIGSILGNDPNSANFQGVRI DAVDNDADLLQIASDYFKEKYRVADNEANAIAHLSIL	595
QY	475	BAWSGNDNDYVDQNNFSLSDINDORSGLMKAFGVASAYRGNLSNLTATAGLKVRSNPDS	534
Db	596	BAWSYNDHQYNKDTKGAQLSDINDPLRETLTTLTFLKSNYRGLSERVITNSLNNRSSEQH	655

Db 165 DQGNVKNFAVSGDKIYFFD-ETGAYKDTSEKVDADKSSSAVSQNA-TIFPAANRAYST 222
QY 217 EASSFETVDNYLTADSWYPRKILKNQSQWASSEGDLRPILMTWMPDAATKAAYANFWA 276
Db 223 SAKNFEAVNYLTADSWYPRKILKDKTWTESGKDDFRPLLMANWPDTETKRNVVYMN 282
QY 277 KEGLISGSY--RQNSANLDAATONIQSAIEKKIASSEGNTWMLRDKMSQFVKSQNWSIAS 334
Db 283 KVGVIDKTYTAETSQADLTAAAEVLQARIEQKITSENNTKWLREIASAFVKTOPOMNGES 342
QY 335 ENETVVPNODHMOGGALLFSNSKD--TEHANSDWRLNRPNPTQTGK-QKYFTTN-----YA 388
Db 343 EK-----PYDDHLONGALLFDNQDTLTPDQTSYRLLNRPNTQTGSLDSRFTYNDPLG 398
QY 389 GYELLANDVNSNPVQAEQLNHLHYLMNWGDIVMGDKDANFDGVYRVDADVNDVADNADLLQ 448
Db 399 GYDFLLANDVNSNPVQAEQLNHLHYLLNFGSIYANDADANFDSIRVDANEDNVDADLQ 458
QY 449 IQBDYYKAKYCTDQNEKNAIDHLSILEANGSNDNDYVKDQNNFSLSIDNDQSGMLKAPG 508
Db 459 ISGDLKAAAYGIDKNNKNANNHVSIVEAWSNDPTPYLHDDGDNLMNDKNFRLSLMLSIA 518
QY 509 YASAVRGNLSNLATAGLKNRSANP-DSDPVPNVFIRAHDSVQTRIAKIIREKLGKTN 567
Db 519 KPTDVRSGLNPLHNSLVREVDREVEVPSYSFARAHDSVQDIIRIIKAEI-NPNS 577
QY 568 DGLTNLTDLDLNKAADIYNQDMNATDKVYYPNNLPMAYMMLQNKDVTVRVYGYDMYDN 627
Db 578 FGY-SPTQEEIDQAFKIYNEDLKSKDKKTHYVPLSYTLTLTNKSGIIPRVYGYDMFTDD 636
QY 628 GQYMATKTPFYNAIETLLKGRIKYVAGGOAVSKQDWSSGILTSVRYGKANSASDAGNT 687
Db 637 GQYMANKTNYDAIESLLKARKMYVAGGOAMQYQIGNGEILTSVRYGKALKQSDKGA 696
QY 688 ETRNSGALLINRPNFRA-YRNLTLNMGAAHKSQAYRPLLLSTKDGIAITYLNDSDVD-S 745
Db 697 TTRTSGVGMGNPFLDGKVVALLMGAAHANQBYRALMSTKDGVIATYATDADASKA 756
QY 746 ROYKYTDSQGNLSFSASHELQSVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKGNLY 805
Db 757 GLVKRTDENGILYFLNDDLKGVANPQVSGFLQVWVPVGAADQDQIRVAASDASTADGKSL 816
QY 806 HQSDALDSQVIEYGFNSFOAFAQSPQYTNVATKNGDLFKSNGITQFEMAPQYVSSDG 865
Db 817 HQDAAMDSDRVMPFGFNSFQSFATKEEBEYTNVIANNVDFKVSNGITDFEMAPQYVSSD 876
QY 866 TFLDSVILNAGYAPSDRYDLAMSNNKYGSKQDLANAIKGLQAGIKVLSDLPVNLNLP 925
Db 877 QFQDSVIQNGYAFDTRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWPDQMYTFP 936
QY 926 GREVVATRVNOYQOAKSGATINKTPYVANTRSYG-DYQEQYGGKFLDLQKLPLPLFST 984
Db 937 KQEVVTVTRDKFGKPIAGSQINHSLYVDTKSSGDDYQAKYGGAFDELKEKYPPELFTK 996
QY 985 KQISTGKPIDPSVKIINWAKYFNGSNILGRGKAYVLE--GNKYLNLAGDKLFLPTVLN 1042
Db 997 KOMSTGQAIDPSVKIKQWAKYFNGSNILGRGADYVLSQVSNKYFNVAASDTLFLPSSLL 1056
QY 1043 NTYQOPQVSANGFISKNGGHHYLDKNGQEVKGRFKEISGSWYVFDSDGKMGATKTKIGND 1102
Db 1057 GKVESGIRYDG-----KGYIYNSSATGDQVKASFITEAGNLVYFGKDGVMVTAQTINGA 1112
QY 1103 TYLFMPNGKQKGVWYDGK-KAYYYDDNGRTWNK-GFVEFRVDGQDKRWYFNGDGTTA 1160
Db 1113 NYFFLENGTALRNTIYTDAGNSHYANDKRYENENGYQQF---GND-WRYFK-DGNWA 1167
QY 1161 IGLVSLD----- 1167
Db 1168 VGLTVDGNVQVFPDKGVQAKDKIIIVTRDGKRVYFDQHNNGNAVNTFIADKTHGYWYLGK 1227
QY 1168 -----NRTLYFPAYGQVKGQVTTIN-GKSYTFDADQDGL----- 1201

Db 1228 DGVAVTGAQTVGKQKLYFEANGQQVKQDFVTSDEGLKLYFYDVSDDGMWMTDTFIEDKAGNW 1287
QY 1202 ----- 1201
Db 1288 FYLGKDGAAVTGAQTIRGOKLYFKANGQOVKGDIVKGTGDKIRYYDAKSGBGEQVFNKTVKA 1347
QY 1202 -----VOTDNANPAP-----OQO-----AGWKLGLGDNOMGYRKD 1230
Db 1348 ADGKTYVIGNDQVAVDPVSVKQTFKDGASGALRFNVLGQLVTGSGWVETANHDWYIQS 1407
QY 1231 GOLLTGEOTIDQKQVFFQDNGVQVKGGTATDASGVLRFYDRDQGHQ----- 1276
Db 1408 GKALTGEQTINGOHLVFKKGHQVKGLQVLTGTGDKRVYYDANSBGDAFNKSVTVNKGTTY 1467
QY 1277 -----VGKGYSTSDNNWYVNESGOVLTL 1302
Db 1468 FQNDGTAQTAGNPKGQTFKDGSDIRFYSMEGQLVTGSGWYSNAQGWLYV-KNGKVLTL 1526
QY 1303 OTIDGQTVYFDDKGIQAKGKAVWDENGLRYFDDADSGNMLDRWKNVGNWYTFNRNG 1360
Db 1527 QTVGSRVYFDENGIOAKGKAVRTSDGKIRYFDENSGSMITNQWKEVNGRYYVFGNDG 1584
RESULT 13
GTFC_STRMU
ID GTFC_STRMU STANDARD: PRT: 1455 AA
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).
GN Name-gtfC; OrderedLocusNames=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
RN NCBI_TaxID=1309;
RN [1] NUCLEOTIDE SEQUENCE.
RP STRAIN=GS-5;
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;
RX Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RX Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RX Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RX Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RX Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RX Qian Y., Li S., Zhu H., Najjar P.Z., Lai H., White J., Roe B.A.,
RX Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 1-349.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RX Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to

CC play a key role in the development of the dental plaque because of
 CC their ability to adhere to smooth surfaces and mediate the
 CC aggregation of bacterial cells and food debris.
 CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
 CC fructose + (1,6-alpha-D-glucosyl)(n+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
 CC forms of glucans.
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -1- SIMILARITY: Contains 5 cell wall binding repeats.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC ENBL; M2054; AAA88592.1; -; Genomic DNA.
 CC ENBL; D88652; BAA26102.1; -; Genomic DNA.
 CC ENBL; D88655; BAA26106.1; -; Genomic DNA.
 CC ENBL; D88658; BAA26110.1; -; Genomic DNA.
 CC ENBL; D88661; BAA26114.1; -; Genomic DNA.
 CC ENBL; D88663; BAA26120.1; -; Genomic DNA.
 CC ENBL; D88665; BAA26120.1; -; Genomic DNA.
 CC ENBL; ABO14940; AAN58706.1; -; Genomic DNA.
 CC ENBL; M17361; AAA88589.1; -; Genomic DNA.
 CC PIR; J0345; J0345.
 CC HSP; P06653; I48G.
 CC InterPro: IPR002479; Cell wall bd put.
 CC InterPro: IPR003318; Glyco_hydro_70.
 CC Pfam; PF01473; CW_binding_1; 2.
 CC Pfam; PF02324; Glyco_hydro_70; 1.
 CC Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
 CC Transference.
 CC -----
 CC SIGNAL 1 34
 CC CHAIN 35 1455 Glucosyltransferase-SI.
 CC REPEAT 1126 1159 A repeat.
 CC REPEAT 1169 1200 A repeat.
 CC REPEAT 1227 1238 C repeat.
 CC REPEAT 1253 1303 AC repeat.
 CC REPEAT 1316 1330 A repeat (incomplete).
 CC REGION 335 1050 Catalytic (approximate).
 CC REGION 1126 1455 2.4 A, 1 C and 1 AC repeats.
 CC REGION 1126 1455 Glucan-binding (approximate).
 CC VARIANT 21 21 V -> I (in strain GS-5).
 CC VARIANT 81 81 P -> L (in strain MT4239).
 CC VARIANT 106 106 D -> V (in strain GS-5).
 CC VARIANT 116 116 S -> A (in strain GS-5 and strain
 CC MT4467).
 CC VARIANT 126 126 A -> T (in strain GS-5).
 CC VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239
 CC and strain MT4467).
 CC VARIANT 256 256 A -> V (in strain GS-5 and strain
 CC MT4467).
 CC VARIANT 425 425 R -> N (in strain MT4251).
 CC VARIANT 519 519 Y -> D (in strain MT4245 and strain
 CC MT4251).
 CC VARIANT 538 538 R -> K (in strain MT4245 and strain
 CC MT4251).
 CC VARIANT 545 545 Y -> F (in strain MT4245 and strain
 CC MT4251).
 CC VARIANT 597 597 N -> D (in strain MT4245, strain MT4251,
 CC strain MT4467 and strain MT8148).
 CC VARIANT 600 600 R -> K (in strain MT4245, strain MT4251,
 CC strain MT4467 and strain MT8148).
 CC VARIANT 601 601 A -> T (in strain GS-5).
 CC VARIANT 614 614 M -> T (in strain GS-5).
 CC VARIANT 727 727 T -> I (in strain MT8148).
 CC VARIANT 734 734 A -> V (in strain MT8148).
 CC VARIANT 964 964 L -> F (in strain MT4239).
 CC VARIANT 1113 1113 N -> Y (in strain MT4239).
 CC VARIANT 1118 1118 A -> T (in strain MT4239).

FT VARIANT 1204 1204 I -> V (in strain GS-5, strain MT4239,
 FT strain MT4467 and strain MT8148).
 FT VARIANT 1208 1208 V -> I (in strain MT8148).
 FT VARIANT 1292 1294 DGH -> NGY (in strain GS-5, strain MT4467
 FT and strain MT8148).
 FT VARIANT 1305 1369 Missing (in strain MT4245).
 FT VARIANT 1326 1326 I -> V (in strain MT8148).
 FT VARIANT 1331 1331 T -> A (in strain GS-5, strain MT4239,
 FT strain MT4467 and strain MT8148).
 FT VARIANT 1377 1377 R -> K (in strain MT8148).
 FT VARIANT 1398 1398 V -> I (in strain MT8148).
 FT VARIANT 1424 1424 D -> N (in strain MT4239).
 FT VARIANT 1439 1439 V -> I (in strain MT4239 and strain
 FT MT8148).
 FT VARIANT 1444 1444 S -> P (in strain MT8148).
 FT CONFLICT 1337 1455 ORLYFKNVQAGELITKRGRIKYKYPDNGNEVRNRYVR
 FT TSSGWNVYFGNDGVALIGHVVEGRVYDENGVEYRASHD
 FT QRNHWYDYYRRDPRGSSASAFRRHSRNGFFDNFRF ->
 FT HASILSLMVFRLRLESSLQSVKVVSTNTMLIPKMFVIM
 FT (in Ref. 1).
 SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A99A4FEC86 CRC64;
 Query Match 41.8%; Score 3020; DB 1; Length 1455;
 Best Local Similarity 45.8%; Pred. No. 1.2e-139;
 Matches 657; Conservative 223; Mismatches 441; Indels 112; Gaps 33;
 QY 1 MEKNRYKLHKVKOWVAIGVTT--VTLGSLAGQVVAADTNNNDGTSQVQNMVPSDPK 58
 DB 1 MEKVRFKLKVKRWTVSVASAVVTLTSL--GSLVKADSTDROQAVT----- 49
 QY 59 FDAQONGLOAQAMFKAANOAD-QTATSQVSPATDGRVDN-QVTPAANOFAANVANQDVA 116
 DB 50 -ESQASLVTTSEAAKETLATDTSTATSQPTATVTDNVSTTQSTNTTANTANFDV- 107
 QY 117 NPATDAGALNROSADTS-----TDGKAVPQTSD----- 145
 DB 108 KPTT-----TSEQSKTDNSDKIIATSKAVNRLTATGKFPVANNNTAHSRTVTDKIVPKPK 163
 QY 146 -----QPG-----HLETVDGKTVYVDANGQRLKNYSWIDGKTVYFQDQTG 186
 DB 164 IGKLPQPSLSQDDAALGNVKNIRKNGKYKYYKEDGTLOKYNALNINGKTFPFBETGA 223
 QY 187 EAQTDLF-KTGOANQNDPDSYQAOANQAYSNEASSGFETVDNYLTADSWYRPRKILKNGS 245
 DB 224 LSNNTLPKKGNI-TNNNDNTNSPAQYQVYSTDAANFEHVDHYLTABSWYRPKYLDGKT 283
 QY 246 WQASSEGDLRPLMTWPDAAATKAAAYANFWAKEGLISSYRQNSA--NLDAATQNTQSAI 303
 DB 284 MTQSTEKDFRPLMTWPDQETQRQYVNTMNAQLGIHQTYNTATSPQLNLAAQTITQKI 343
 QY 304 EKKIASGNTNRLDKMSQFVKSQNWSIASENETVYPNODHMOGALLPS-NSKDEHA 362
 DB 344 EKITAETKNTNWLRTISAFVKTSANWSDSK----PFDDHLQKGLALYSNNKSLTQA 399
 QY 363 NSDMRLNRNPTFQTGKQ--KYFT--TNYAGYELLANDVNSNPVQAEQLNHLHYLMNW 419
 DB 400 NSNYRILNRNPTNQTGKDPRTADRTIGGYEFLLANDVNSNPVQAEQLNHLFLMNF 459
 QY 420 GIVMGDKDANPDGVRVDADVNNADLLQIQDYKAKYGTQONEKNAIDHLISLEAWSG 479
 DB 460 GNIYANDPDANFDSIRVDADVNDADLLQIAGDYLKAAGIKHNKDKAANDHLSLEAWSY 519
 QY 480 NNDVYKQNNFSLISIDNDQSRGMLKAFGYASAYRGNLSNLATAGLKNS-ANPDSDPVP 538
 DB 520 NDTPLYHDDGDMNINMDNRLRLSLYSIAKPLNQSRGMPLITNSLVNRTDNDNAEAAVP 579
 QY 539 NYVFIKRAHDSVQTRIAKIREKLGKTNADGLTNLTLDLNLKAFDITYNODMNAATKVYYP 598
 DB 580 SYSFIRAHDSVQDILIRNIIRAEI-PPNVVGY-STMEIKKAFYINXKDLATEKKYTH 637
 QY 599 NNLPMAYAMMLQNKDVTTRVYVGDVMTNGQVMAKTTPPYNAIETLLKGRIKYVAGQAV 658
 DB 638 YNTALSYALLTNTKNSVPRVYVYGDVMTDQGMMAHTINYEAETILLKARIKYVSGQAM 697

QY 659 SYKQDSSGILTSVRYGKGANSAGNTETRTNSGALLINRPNFR--AYRNLTLNMGCA 716
Db 698 RHQQVGNSEIIYSVRYGKALKATDGTDRRTSGVAVIEGNNPSRLKASDRVVNMGCA 757
QY 717 AHKSQAYRPLLLSTKGIATYLNDSVDSDRQKYTDSQGNLSFSASELSQSVANAQVSGMI 776
Db 758 AHKQAYRPLLLTTDNGIKAYHSDQEA-AGLVRYTNDRGELIFTAADIKGYANPQVSGYL 816
QY 777 QWVPVGAADNODVRSPTQATKQNIYHQSDALDSQVYGGFNSFQAFQPSOYTNVA 836
Db 817 GVMVPVGAADQDVRVAASTAPSTDGKS VHQNAALDSRVMFEGFSNFQAFATKCEYTNV 876
QY 837 VTAKNGDLKPSGWIITOFENAPQVSSDGTFLDSVILANGYAPSDRYDLAKMSKNKYGSKQ 896
Db 877 VTAKNVKFAEAGVTDVFENAPQVSSDGTGFLDSVILQNGYAPTDRYDLGSKENKYGTAD 936
QY 897 DLANAIKGLQAGIKVLSDLPVNLNPKGVWVATRVNOYGOAKSGATINKTPVANT 956
Db 937 DLVKAIALHSGIKVMADWPDQWYALPEKEVWVATRVDKYGTPTVAGSQIKNTLYVVDG 996
QY 957 RSYG-DYQSYGKGLDDLOKLYPLRFLSTQKISTGKPIDPSVKITNWSAKYFNGSNILGR 1015
Db 997 KSGSKDQQAQYCGAFLEELQAKYPELFARKQISTGVPMDEPSVKIKOWSAKYFNGTNILGR 1056
QY 1016 GAKYVLSLSE--GNKYLNLADGKFLPTVILANTYGOQPOVSANGFISKNGGHIYLDKNGQEVK 1073
Db 1057 GAGYVLKQDQATNTYFSLVSDNTFLPKSLVNPNGHTSSSVTGLVFDGKGVYVYSTGNOAK 1116
QY 1074 NRPKEISGWSYFDSCKMATGKTKIGNDTYLFMPNGKOLKGVWYDGKKA-YYDDNGR 1132
Db 1117 NAFISLGNWYFDPNNGYVMTGAQSLNGANYFPLSNGIQIRNAIYDNGKVLSYGNDGR 1176
QY 1133 TWTNKGFBVRVDGQDKRYFNGDGTIAIGLVSLEDRNTLYFDAYGYQVKGQTV-TINGKS 1191
Db 1177 RYEN-GYLLF--GQ-QWRYFO-NGIMAVGLTRIHCAGVQYFDASGFOAKGQFITADGKL 1230
QY 1192 YTFDADQGLVQTDNANPAQOAG-WKLLGDNGQYKXGQGLLTGEQITDQKVPFQDN 1250
Db 1231 RYFDRDSGNOI-----SNRRVNRNSKGEWFLFDHN-----GVAVTGTVTFNGQRLYFKN 1279
QY 1251 GVOVKGTATDASGLRFDYRQGHQVKGWSTSDDNWVYNSOVLTLQTTDQGTV 1310
Db 1280 GVOAKGEFTDADGHLRYDPSNGEVNRVFRNSKGEWFLFDHNGIATGTRVNGQRL 1339
QY 1311 YFDDKGIQAGKAVMDENGLRYFDADSGNMLDRW-KNVVDGNWYFNRNGLA 1362
Db 1340 YFKSNGVQAKGELITERKGRIKYDPSNGEVNRVYRTSSGNYFYFGNDGVA 1392

RESULT 14

Q55263 9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GTF-I.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=131d;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase
RT produced from Streptococcus sobrinus ATCC 33478.";
RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).
RN [1]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;

RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D63570; BAA09792.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; 1GVW.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MM; 9DF7A3F2C6B4FD43 CRC64;

Query Match 41.7%; Score 3018; DB 2; Length 1590;
Best Local Similarity 41.6%; Pred. No. 1.8e-139;
Matches 674; Conservative 210; Mismatches 440; Indels 298; Gaps 38;

QY 1 MEKRLRYLKHVKKQWVAIGVTVTVT-LSFLAGGVVVAADTNNNDGTSVQVNVKWFSDPKF 59
Db 1 MEKNVRFKQHKVKKRWVTLVSASATMLASALGASVASADTDTASDDSNQA--VVTGD--- 55
QY 60 DAQAQNGQLAQAMPKAAQADQATSTQVSPATDGRVDNQVTPAANQPAANVANQDV---- 115
Db 56 --QTNNQATDQTSIAA-----TATSEQASASTDAATDQ--ASAEQQTQGTASTDTAAQT 106
QY 116 ---ANPA-----TDAGALNRQSAADTSTDGKAVPQTSQDQGHLETVDGKTYTV 160
Db 107 TTNAEAKWPTENENOGFTDEMLAEAKNVATAESD--SFPSDLAKMSNVKQVDGKYVY 164
QY 161 DANGQRLKNYSNVIDGKTYVFDGQTG---BAQTDLPKTGOANQDNVPSYCANNOAYSN 216
Db 165 DQDGNVKNFAVSGVDKIYFD-ETGAYKDTSKVDADKSSSAVSQNA-TIFAANNRAYST 222
QY 217 EASSPTVDNYLTADSWYRPRKILKNGSQWASSEGDLRPIILMTWMPDAATAKAAVANFWA 276
Db 223 SNAENFADVNYLTADSWYRPRKILKNGKTWESGKDDFRPLJMANWPDTEKGYVNYM- 281
QY 277 KEGLISG-----SYRQNSANILDAATQNTQSAIEKKIASEGNTNWLDRKMSQFVKSQNWS 331
Db 282 --NLVVGIDKTYTAETSQADLTAAAEVQARIEQKITTEQNTKWLREAIASFVKTPQWN 339
QY 332 IASENETYVPQDHNQGGALLFSNSKD-TEHANSDWRLNRRNPTQTKG-QKYFTTN--- 386
Db 340 GESEK-----PYDDHLQNGALKFPDQSDLTPTDQSNYRLNRRPTQTKGSLDSRFTYAND 395
QY 387 -YAGYELLANDVDNSNFPVQAEQLNHLHLYLNNWGDIVMGDKDANFQDGVVRVDADVNNAD 445
Db 396 PLGGYEFILLANDVDNSNFPVQAEQLNHLHLYLNFQSIYAKDADANFDSIRVDAVDNDAD 455
QY 446 LLQIQORDYYKAYGTDQNEKNAIDHLSILEAWSGNDNDYVQDQNNFSLSIDNQSGMLK 505
Db 456 LLQISSDYLKAAAGIDKKNKNNHVSIVEAWSNDTPLYLHDDGNLNMNDKFRLSMLW 515
QY 506 AFGYASAYRGNLSNLATAGLKNSRANP-DSDDPVNVPVFIKRAHDSVQVFIKRIKELK 564
Db 516 SLAKPLDKRSGNLNPLHNSLVDREYDREVETVPSYSFARAHDSVQDILIRDIKAEI-N 574
QY 565 TNADGLTNLLTDLNKAFDIYNQDMMATDKVYYPNNLPMAYAMWLQNKDVTVRVYGYDMY 624
Db 575 PMSFGY-SFTQEEIQAPKIYNEDEKKTDDKKTHTNVPLSYTLTLTKNGSIPRVYGDNF 633
QY 625 TDNGQYMATKTPFYNAIETLLKGRIKYVAGGQAVSYKQDWSGGIITSVRYGKANSASDA 684
Db 634 TDDQYMWANKTVNYDAIESLLKARMKYVSGQAMQNYQINGEILITSVRYGKALKQSDK 693
QY 685 GNTETFRSGMALLINRRPNFRA-YRNLTLNWCAAHKSQAYRPLLLSTKDGITATYLNDSV 743
Db 694 GDATTRTSGVGMVQNGQPNFSLDGKGVVALNMGAAHANOEYRALMVSTKDGVTATYADADA 753
QY 744 D-SROQKYTDSQGNLSFSASELSQSVANAQVSGMIQWVPVGAADNQDVRTSPSTQATKDG 802

Db 754 SKAGLVKRTDENGVLFLNDLKGAVNPQVSGFLQVWPVGAADDODIRVAASDTASTDG 813
QY 803 NIYHQSDALDSQVIYEGFSNFQAFQSPDOYTNVIAKNGDLFKSGGITOFEMAPQVSS 862
Db 814 KSLHQDAANDSRVMFEGFSNFQSPATKEEYTNVIANVDFKFSWGITDFEMAPQVSS 873
QY 863 EDGTFDLSVTLNGYASDRYDLAWSKKNYKGSQDLANAIKGIQSAGIKVLSLDLVPNQLY 922
Db 874 TDGQFLDSVTLNGYAFETDRYDLGMSKANKYGTADQLVKAIKALHAKGLKVMADWVPQMY 933
QY 923 NLPKGEVWATRVNOYGOAKSGATINKTPVAVNTRSYG-DYQYQYQYKGLDDIQLKLPRL 981
Db 934 TFPKQEVVTVTRDKFGKPTAGSQINHSLYVTUTKSGDDYQAKYGAFLDELKEYPEL 993
QY 982 FSTKQISTGKPIDSPSKITNWSAKYFNGSNILGRGAKYVLSE--GNKYLNLAADGKFLP- 1038
Db 994 FTKKQISTGAIDPSVKIKQMSAKYFNGSNILGRGADYVLSQVSNKYFNVASDTFLPLS 1053
QY 1039 -----TVLNNNT-----YQOPQVSAN-----GPIKNG 1060
Db 1114 YPFLNGTALRNTIYTDAGNSHYANDGKRYENGYQQFGNDWRYFKQGNMAYGLTTVDG 1173
QY 1061 GIHYLDKNGQEVKNR-----FKETSGSWYTFDSGKMATGK 1096
Db 1174 NVQYFDKGVQAKDKIIVTRDGKRVYFDQHGNAVNTFTADTKGHHYILGKDKVATGA 1233
QY 1097 TKIGNDTYLFMPNGKOLK-----EGWYDGGKAYYD-DNGRTWTKNGFVEFRVDGQD 1148
Db 1234 QTVGKQKLYFEANGEQVKGDVFTSHEG-----KLYFYDVSDDMTDT-FTB---DKAG 1283
QY 1149 KRYFNGDGTIALGLVSLDNRTLYFDAYGVQVKGQVTV-INGKSYTFDADQGLVQTDNA 1207
Db 1284 NMFYLGKOGAAVSGAQTIRGQKLYFKAYGOQVKGDIVKGTGDKIRYIDAKSGEQVFNKT 1343
QY 1208 NPAP-----QGO-----AGWKLLEDGNDQWYR 1228
Db 1344 KAADGKTYVIGNNAVDPVSVKQTFKDGASGALRPNLKGQVTCGHWYETANHDDWVYI 1403
QY 1229 KDQQLLTGQTIDGQKVPFQDNGVQVKGGTATDASGLVRFYDRDQGHQ-----1276
Db 1404 QSGKALTGQTIDGQKVPFQDNGVQVKGGTATDASGLVRFYDRDQGHQ-----1276
QY 1277 -----VKGWYTSDDNNWYVNESQVLT 1300
Db 1464 YFPGNDGTAQTAGNPKGIQFKDGSVLRFYSMEGQLVIGSGWYSNAQGOVLYV-KNGKVL 1522
QY 1301 GLQTTIDGQTVYFDDKGIQAKGAVDENGRLRYFDADSGNMLRDRKWKYDGNWYFNRNG 1360
Db 1523 GLQTVGSQRVYFDENGIQAKGAVRSDGKIRYFDENSGSMITNQWKFYQYIYFNGDG 1582
QY 1361 LA 1362
Db 1583 AA 1584

RESULT 15
Q9WXAJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WXAJ5_9STRE
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE GTF-S.
GN Name=gtf;
OS Streptococcus criceti.
OG Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;

RN [1]
RP NUCLEOTIDE SEQUENCE.

RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSSP; P06654; IMPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 41.4%; Score 2992.5; DB 2; Length 1512;
Best Local Similarity 46.3%; Pred. No. 2.9e-138;
Matches 663; Conservative 214; Mismatches 452; Indels 107; Gaps 40;

QY 1 MEKNRYKLHKVKQWVAIGVTTVLSFLAG----GQVVAADTNNN-----DGTSVQV 49
Db 1 MEKNRYKLHKVKQWVAIGVTTVLSFLAG----GQVVAADTNNN-----DGTSVQV 49
QY 50 NK-MVPSD--PKFDAQONGQLAQAMFKAAQADQATATQSVSPATDG-RVDNQ-VTPAAN 104
Db 61 NKEVDSDAQAATDAKATSEQAASVSDTNSSETDQVN-----VDGVSANNQPEAPAAP 114
QY 105 QPAA-----NVANQDVANPATDA-----GALNR--QSAADTSTDGKAVPQTSDPQHLETV 153
Db 115 QAAASNTNTATSEANTNTAVSEAPAAENRTAEKGADLSQEEAALSLD---NIKKV 171
QY 154 DGKTYVVDANGQRLKNYSMVIDGKTYFDGQGTGEAOTDLPKTQGAQNDVNP--DSYQANN 211
Db 172 DKYTYVVDANGQRLKNYSMVIDGKTYFDGQGTGEAOTDLPKTQGAQNDVNP--DSYQANN 211
QY 212 QAYSNEASFETVDNYLTADSWYRPRKILKNGSQWASSEGDLRPLMTWPDAAATKAAAY 271
Db 230 KAPDSSAKSFELVDGYLTAEASWYRPTKILENGKTWVDSKETDLRPLVTLSWPKDTQVAY 289
QY 272 ANFWAKE--GLISGSYRQNSANLDAQNIQSAIEKKIASEGNTWLRDKMSQFVKSONQ 329
Db 290 LNYMSKALGQKEEFTTKTSQTALNTAAEMIQMKIEQISKEQGTAWLRDAMAFAVATQSR 349
QY 330 WSIASENETVYVNOHMOGQALLFNSKDETHANSWRLNLRNPTFQTKQKTYFTTN-VA 388
Db 350 WNSDSEQ---FKNHDLQGGALLYNNKLTETWADSKYRLNTRPTPDQDKTHYSKADYEG 406
QY 389 GYELLANDVDNSNPVQAEQLNHLHLYLMNWGDI VMGDKDANFDGVRVDAVDNVDNADLLQ 448
Db 407 GYEFLLANDVDNSNPVQAEQLNHLHLYLMNWGDI VMGDKDANFDGVRVDAVDNVDNADLLQ 448
QY 449 IQRDYKAKYGTQDQNEKNAIDHLSLEAWSGNDNDYVDQNNFSLISINDQDSGMLKAFG 508
Db 467 LYTTFNAYVYGVKSEQAQALAHISILEAWSYNDNYNDQNTGAALMDNGLSLLYTLT 526
QY 509 YASAYR-GNLSNLATA--GLKNRSANPD--SDPVNYPVETRAHDSVQTRIAREKLK 564
Db 527 RPLSERTPLGLSTLIKSEYGLTDRTKDDYGDTPQSPYFVRHSDSEVQTVIAIKKEKIDP 586
QY 565 TNADGLTNLTDLNLKAFDIYNQDNNAATDKVYPPNNLPMAAYAMLQNKDQTVTRVYVYGD 624
Db 587 T-TDGTG-FTLDQLAQAFDIYNKDMNSVEKHYTHYNIPAAVAVMLSNMESVTRVYVYGD 644
QY 625 TDNGQYMATKTPFYNAIETLLKGRJYVAGGQAVSYK-----QWSSG-----667
Db 645 TDDGQYMETKSPYDAINTLLRARIYAAGGQTMESHKAYTPSAAMKAKPDSPGSLVNSG 704
QY 668 ILTSVRYGKANSASD-AGNETRNSGMALLINNPFR--AYRNLTANWGAHKSQAVR 724
Db 705 VLVSVRFQDVMASDDMTGGQLAKTSGMPSLLANPELDAEELIKVNVGKIHAGQAVR 764
QY 725 PLLLSTKDGATYLNDSVDVSRQYKYTDSQGNLSFSASELQSVANAQVGMIVQVVPVGA 784

Db 765 PLLLTDTKGLQKYLNDSDTNLT--KVADKDGFTFKGSEIKGYKQVEVNGYLSVWVPVGA 822

Qy 785 ANQDVRTSPSTOATKQGN-1YHQSDALDSQVIYEGFSNFQAPQSPDQYTNAVIAKNGD 843

Db 823 KSDQIRVAASTKANAKGDKSVTASQALDSQLIYEGFSNFQDFVQKDAQYTNKKIAENTD 882

Qy 844 LFKSWGITOFEMAPQVYVSSDGTFLDSVLNGLYAFSDRYDLAMSKNKKYGSQODLANAIK 903

Db 883 LFKANGVTSFEMAPQVVSATDGTFLDSIIQNGYAFSDRYDLAMSKNKKYGSQKEDLANALK 942

Qy 904 QLOSAGIKVLSLVPNOLYNLPCKEWTATRVNOYGOAKSGATINKTPYVANTRSYG-DY 962

Db 943 ALHAGIOAIAWVPDQIYQLPCKEWTATRVNOYGRVKIDQPMWNKLYLANTKSGKDF 1002

Qy 963 QBQYGGKFLDDIQLKLYPLRFLSTKQISTGKPIDPSVKITNWSAKYFNGSNILGRGAKYVLS 1022

Db 1003 QAKYGGFLAEIQLQYPEMFTAKMISTGKPIDSSVKLKWSAQYFNGTIVLGRGTDYVLS 1062

Qy 1023 -EG-NKYLNLADGKLFPLTVLANTYQPOQVSANGFISKNGGIHYLDKNQOEVKNRPEKIS 1080

Db 1063 DEGTGKYFTVNEKGEFLPAVLTG----DKEAKTGFYNDGKGMTYFTTAGSQAKSDFVTVA 1118

Qy 1081 GSWYFYFSDGKMATKTKIGNDTYLFMPNGKOLKEGVWYDGK-KAYYDDNGRTW---T 1135

Db 1119 GNTYFYFDYTGHWVTGPNGINTKPYFPLFNGVMLKDAVMEDDRGRSVYKGTGMKGSRN 1178

Qy 1136 NKGFVEFRVDGQDKWRYFNGDGTIAIGLVSLDNRTLYPDAGYQVKGQTVTTIN-GKSYTF 1194

Db 1179 NEWFAMTDSKGQLRFRHPDNYGFMSVGLVTIHGNVQYDEEGFQVKGDFTDKAGQTRYF 1238

Qy 1195 DADQGLVQTDNANPAPQOQACWKLLGDNQWGYRKD-GOLLTGEQTIIDQKVFFQ-DNGV 1252

Db 1239 DXNTGNLVKGQPFNQ-----NGHWYYSDDQGLIAKGAQTIKGQKLYFDAKTGA 1286

Qy 1253 QVKGGTATDASGLRFRYDRDQGHQVGKGYSTSDNNWVTVNESGOVLTLQTIIDGQTVYF 1312

Db 1287 QVKGDFVTDKGNWTFYSGTGDGLAVSTFFSTGNNAWFYADENGHVAKEKTINGQKLYF 1346

Qy 1313 DDK-GIOAKGKAVMDENGNLRYFDADSGNMLRDRW---KNVDGNWYFFNRRGLATR 1364

Db 1347 DTKGQOAKGRFVRDAKG-LRFYDADTGALVTNSFLETKAGSNQWYIMGADGYAVR 1401

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 151.097 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-40
Perfect score: 7928
Sequence: 1 MENKHYKLHKVKQWTTIA.....PGVYVFDKNGLAYPRVLN 1518

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues
Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*
9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7928	100.0	1518	7	Add93660 Streptoco
2	7928	100.0	1518	9	Adx37283 Streptoco
3	3818	48.2	1554	7	Add93658 Streptoco
4	3818	48.2	1554	9	Adx37281 Streptoco
5	3514.5	44.3	1430	5	Aau98044 S. mutans
6	3510.5	44.3	1430	5	Aau98045 S. mutans
7	3509.5	44.3	1430	5	Aau98041 S. mutans
8	3507.5	44.2	1430	5	Aau98043 S. mutans
9	3505.5	44.2	1430	5	Aau98042 S. mutans
10	3502.5	44.2	1430	5	Aau98029 S. mutans
11	3502.5	44.2	1430	9	Adx37279 Streptoco
12	3502.5	44.2	1430	7	Adx37279 Streptoco
13	3325.5	41.9	1590	7	Add93657 Streptoco
14	3325.5	41.9	1590	9	Adx37280 Streptoco
15	3296.5	41.6	1592	2	Aar32925 Glucosylt
16	3253	41.0	1476	5	Aau79284 Streptoco
17	3196.5	40.3	1475	5	Aau98027 S. mutans
18	3196.5	40.3	1475	7	Add93654 Streptoco
19	3196.5	40.3	1475	9	Adx37277 Streptoco
20	3195.5	40.3	1475	5	Aau98030 S. mutans
21	3192.5	40.3	1475	5	Aau98040 S. mutans
22	3192.5	40.3	1475	5	Aau98031 S. mutans
23	3190.5	40.2	1475	5	Aau98035 S. mutans
24	3190.5	40.2	1475	5	Aau98033 S. mutans

25	3189.5	40.2	1475	5	AAU98032	Aau98032 S. mutans
26	3188.5	40.2	1475	5	AAU98036	Aau98036 S. mutans
27	3184.5	40.2	1475	5	AAU98034	Aau98034 S. mutans
28	3182.5	40.1	1475	5	AAU98037	Aau98037 S. mutans
29	3172.5	40.0	1475	5	AAU98038	Aau98038 S. mutans
30	3169.5	40.0	1475	5	AAU98039	Aau98039 S. mutans
31	3155	39.8	1577	2	AAU91047	Aar91047 Alpha-D-g
32	3070	38.7	1375	5	AAU98028	Aau98028 S. mutans
33	3070	38.7	1375	5	AAU79288	Aau79288 Streptoco
34	3070	38.7	1375	7	Add93655	Add93655 Streptoco
35	3070	38.7	1375	9	Adx37278	Adx37278 Streptoco
36	3062.5	38.6	1499	7	Adc54806	Adc54806 Protein S
37	2883	36.4	2835	5	ABR98574	Abb98574 Dextran s
38	2883	36.4	2835	6	ABR55594	Abb55594 Amino aci
39	2804	35.4	1365	7	Add93659	Add93659 Streptoco
40	2804	35.4	1365	9	Adx37282	Adx37282 Streptoco
41	2795.5	35.3	1497	6	ABR63234	ABr63234 Glucanuc
42	2759.5	34.8	1527	5	Aau80055	Aau80055 Leuconost
43	2750.5	34.7	1527	7	Adc54807	Adc54807 Leuconost
44	2601.5	32.8	1477	9	ADY72732	Ady72732 Mutant de
45	2599.5	32.8	1477	9	ADY72696	Ady72696 Mutant de

ALIGNMENTS

RESULT 1
ADD93660
ID ADD93660 standard; protein; 1518 AA.
XX
AC ADD93660;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus salivarius glucosyltransferase-I.
XX
KW Glucosyltransferase; enzyme; vaccine; anticonvulsant; immunogen.
XX
OS Streptococcus salivarius.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
(FORS-) FORSYTH INST.
XX
Smith DJ, Taubman MA;
WPI; 2003-845091/78.
Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.
Claim 16; Page 16; 49pp; English.

The present sequence is the protein sequence of Streptococcus salivarius glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diabetic or multi-epitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX
PS Claim 7; SEQ ID NO 40; 73pp; English.
XX
CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus salivarius GbpB protein of the invention.
XX
SQ Sequence 1518 AA;
Query Match 100.0%; Score 7928; DB 9; Length 1518;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MENKIHKLHKVKKQWTTIIVASVALATVLGGLSVTTSSVSADETQDKVTQNSGTTAS 60
DB 1 MENKIHKLHKVKKQWTTIIVASVALATVLGGLSVTTSSVSADETQDKVTQNSGTTAS 60
QY 61 LVTSPEATKEADKRTNTEADVLTPAKETNAVETATTTNTQATAEAATATTADVAVA 120
DB 61 LVTSPEATKEADKRTNTEADVLTPAKETNAVETATTTNTQATAEAATATTADVAVA 120
QY 121 PNKEAVTTDAPVTTTEKAEQPAIVKAEVNTVEVKAPEAALKDSEVEAALSUKNTKID 180
DB 121 PNKEAVTTDAPVTTTEKAEQPAIVKAEVNTVEVKAPEAALKDSEVEAALSUKNTKID 180
QY 181 GKYYVNEGSHKENFAITVNGQLLYFGKDGALTSSTSYSTFTPGTTNIVDGFSSINRAYD 240
DB 181 GKYYVNEGSHKENFAITVNGQLLYFGKDGALTSSTSYSTFTPGTTNIVDGFSSINRAYD 240
QY 241 SSEASPELIDGYLTADSWRPAASIIKDGVTWQASTAEDFRPLMAMWPNVDQVNYLNM 300
DB 241 SSEASPELIDGYLTADSWRPAASIIKDGVTWQASTAEDFRPLMAMWPNVDQVNYLNM 300
QY 301 SKVFNLDKYSSTDQKQETLKVAAKDIQIKIEQIKAEKSTQWLRETIISAFVKTQPOWKE 360
DB 301 SKVFNLDKYSSTDQKQETLKVAAKDIQIKIEQIKAEKSTQWLRETIISAFVKTQPOWKE 360
QY 361 TENYSKGGEDHLQGGALLYVNDSTRPWANSYRRINRATNTQGTIDKSIILDEQSDPNH 420
DB 361 TENYSKGGEDHLQGGALLYVNDSTRPWANSYRRINRATNTQGTIDKSIILDEQSDPNH 420
QY 421 MGGFDFLLANDVLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDVNDVADM 480
DB 421 MGGFDFLLANDVLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDVNDVADM 480
QY 481 LQLYTNYFREYGVNKSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKQRLALLFS 540
DB 481 LQLYTNYFREYGVNKSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKQRLALLFS 540
QY 541 LAKPIKERTPAVSPLYNNFTNTQDEKTDWINKDGSKAYNEDGTQKSTIGKYNEKYGD 600
DB 541 LAKPIKERTPAVSPLYNNFTNTQDEKTDWINKDGSKAYNEDGTQKSTIGKYNEKYGD 600
QY 601 ASGNTVFIIRAHNNVDIIIAEIIKKEINPKSDGFTTIDAEKQAPRIYNKMLSSDKYKT 660
DB 601 ASGNTVFIIRAHNNVDIIIAEIIKKEINPKSDGFTTIDAEKQAPRIYNKMLSSDKYKT 660
QY 661 LNNIPAAVAMLQNMETITRVYVYGDLYTDGHHMETKSPYYDTIVNLMKSRIKYVSGGQA 720
DB 661 LNNIPAAVAMLQNMETITRVYVYGDLYTDGHHMETKSPYYDTIVNLMKSRIKYVSGGQA 720
QY 721 QRSYWLPTDGMKDNDSVELYRNEVYTSVRYGKDIIMTANDTEGSKYSRTSGQVTLVANNP 780
DB 721 QRSYWLPTDGMKDNDSVELYRNEVYTSVRYGKDIIMTANDTEGSKYSRTSGQVTLVANNP 780

QY 781 KLNLDQSAKLNVMEMGIHANQKRYALIVGTADGINKFTSDADAIAAGYVKETDSNGVLTFF 840
DB 781 KLNLDQSAKLNVMEMGIHANQKRYALIVGTADGINKFTSDADAIAAGYVKETDSNGVLTFF 840
QY 841 GANDIKGYETFDMSGFVAVWVPVGASDNQDIRVAPSTEAKKEGELTKATEAYDSQLIYE 900
DB 841 GANDIKGYETFDMSGFVAVWVPVGASDNQDIRVAPSTEAKKEGELTKATEAYDSQLIYE 900
QY 901 GFSNFOTIPDGSPPSVYTNRKIAENVDLFSKMGVTSFEMAPQVVSADGDTFLDSVQNGY 960
DB 901 GFSNFOTIPDGSPPSVYTNRKIAENVDLFSKMGVTSFEMAPQVVSADGDTFLDSVQNGY 960
QY 961 AFADRYDLAMSKNKNKYGSKEDLRDALKALHKAGIQAIADWVPDQIYQLPCKEVVYATRTD 1020
DB 961 AFADRYDLAMSKNKNKYGSKEDLRDALKALHKAGIQAIADWVPDQIYQLPCKEVVYATRTD 1020
QY 1021 GAGRKTIADAIIDHSLYLVANSKSSGKDYQAKYGGEFPLAELKAKYPENFKVNMISTGKPIDD 1080
DB 1021 GAGRKTIADAIIDHSLYLVANSKSSGKDYQAKYGGEFPLAELKAKYPENFKVNMISTGKPIDD 1080
QY 1081 SVKLLKQWKAIEYFNGTINVLERGVGYVLSDEATGKYFTVTKEGNFIPLOLTGKEKVIITGFSS 1140
DB 1081 SVKLLKQWKAIEYFNGTINVLERGVGYVLSDEATGKYFTVTKEGNFIPLOLTGKEKVIITGFSS 1140
QY 1141 DGKGIYFCTSGTQAKSAFVTFNGNTIYFDARGHMVTNSEYSPNGKDVYRFLPNGIMLSN 1200
DB 1141 DGKGIYFCTSGTQAKSAFVTFNGNTIYFDARGHMVTNSEYSPNGKDVYRFLPNGIMLSN 1200
QY 1201 APYIDANGNTIYLYNSKGQMYKGYTKFDVSETDKGKESKVVKFPRFTNEGVMAGVTVVI 1260
DB 1201 APYIDANGNTIYLYNSKGQMYKGYTKFDVSETDKGKESKVVKFPRFTNEGVMAGVTVVI 1260
QY 1261 DGFQTVFGSDGFOAKDKLVTFKGTYYFDAHTGNGIKDWTNRNINGKWIYFDANGVAATGA 1320
DB 1261 DGFQTVFGSDGFOAKDKLVTFKGTYYFDAHTGNGIKDWTNRNINGKWIYFDANGVAATGA 1320
QY 1321 QVINGQKLYFNEBDSQVKGKVVKNADGTYSKYKEGFGELVTNEFTTDTGDNVWYIYAGANGK 1380
DB 1321 QVINGQKLYFNEBDSQVKGKVVKNADGTYSKYKEGFGELVTNEFTTDTGDNVWYIYAGANGK 1380
QY 1381 TWTGAQVINGQHLVFNADSGQVKGKVVKNADGTYSKYKYNASTGERLNTNEFTTGDNNWYI 1440
DB 1381 TWTGAQVINGQHLVFNADSGQVKGKVVKNADGTYSKYKYNASTGERLNTNEFTTGDNNWYI 1440
QY 1441 GANGKSVTGEVKIGDDTYFPAKDGKQVKGQTSYAGNGRISYYGDSGKRAVSTWIEIQPG 1500
DB 1441 GANGKSVTGEVKIGDDTYFPAKDGKQVKGQTSYAGNGRISYYGDSGKRAVSTWIEIQPG 1500
QY 1501 VVYVFDKNGLAYPPRVLN 1518
DB 1501 VVYVFDKNGLAYPPRVLN 1518
RESULT 3
ADD93658
ID ADD93658 standard; p:protein; 1554 AA.
XX
AC ADD93658;
XX
XX 29-JAN-2004 (first entry)
XX
DE Streptococcus sobrinus glucosyltransferase-U.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
XX 07-MAR-2003; 2003WO-US006962.
XX

[illegible]

Db 1304 YAENGVYVYKFOVAENQDQWYFYDQTTCKQAKGAQKADGRDLFYFNPDSGVQVKGDFATD 1363
 QY 1410 ADGTVSKYNASTGERLTNEFFTTGDNWYIYGANGKSVTGEVKI----- 1453
 Db 1364 ESGNTSFYHGDNGDKVVGGFTTGNNAWYADNNGNLVKGFQEDIDGKWYHFDEVTGQAK 1423
 QY 1454 -----GDDTYFAKDGKQVKTYSAGNRSYYVYGDGSKRAVSTWETIQGVVYVYDK 1507
 Db 1424 GAALVNGQOOLYFDVDSIGIQVKGDFVTDGQGNYSYYDVNSGDKKXVNGFFTTGDNAMYADG 1483
 QY 1508 NG 1509
 Db 1484 QG 1485

RESULT 5

AAU98044

ID AAU98044 standard; protein; 1430 AA.

XX AAU98044;

AC AAU98044;

DT 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFD mutant N471D/T589D.

XX Glucosyltransferase; GTFD; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; muten.

XX Streptococcus mutans.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 471

FT /note= "wild-type Asn substituted by Asp"
 FT Misc-difference 589

FT /note= "wild-type Thr substituted by Asp"
 FT US2002031826-A1.

PN 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

XX 07-JUN-1995; 95US-00485243.

XX 16-JAN-1998; 98US-00007999.

XX 16-JAN-1998; 98US-00008172.

XX 20-JAN-1998; 98US-00009620.

XX 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful

as substitutes for and additions to modified starch and latexes in paper
 manufacture, comprises mutations in specific positions.

XX Claim 36; Page: 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 B polypeptide having changes at position from I448V, D457N, D567T,
 CC K104T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a

GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the

GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTF mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the Indexer
 CC using the GTFD sequence appearing as AAU98029 and the information in
 CC claim 36

XX Sequence 1430 AA;

SQ

Query Match 44.3%; Score 3514.5; DB 5; Length 1430;

Best Local Similarity 48.2%; Pred. No. 7.1e-180;

Matches 726; Conservative 246; Mismatches 433; Indels 101; Gaps 32;

QY 1 MENKHYKLVKVKQWVTIAVASVALATVGLSLVTTSSVSAD---ETQDKTVTO-SNSG 56

Db 1 METKRRYKMKVKKHVVTVAVAS-GLITL--GTTTLGSSVSAETQQTSKRVYTKQSEDD 57

QY 57 TTASLVTSPKATKADKRTNTKADVLTPAKETNAVETATTTNTQATAEAAAT-ATTADV 115

Db 58 KAASSTQTDAPKTKQAQTEQTAQ-----SQANVADTSTSTTKETPSQNTTQANSDDK 112

QY 116 AVAAVPNKEAVVTDPAVTTKEABEQPATVKAEEVNVTEVKA-----PEALKDSEVBA 169

Db 113 TVTNTKSEEAQTSSE---RTKQSEEAQTASSQAL-TQAKAELTKQRTAAAKNPNVD 167

QY 170 ALSLAKNIKNIDKYYVNVNEDGSHKENFALTVNGQLLYFGKD-GALTSSTSYFTPTTNI 228

Db 168 LAAPVNVKQIDGKYIYIGSDGQPKNFALTVNNKVLVFDKNTGALTDTDSQYQKQGLTKG 227

QY 229 VDFGSINNRADVSSSEAFELIDGYLTADSWYRPASIIKDGVTWQASTADDFRPLMAWVP 288

Db 228 NNDYTPHQIVNFENTSLETIDNYVTADSWYRPKDLKNGKTWTASSESDLRPLMSWVP 287

QY 289 NVDTQVNYLYNYSKV-FNLDAKYSTDDKQETLVKAAKQIQIKIEOKIAEKSTQMLRETI 347

Db 288 DKQTQIAYLVNMQGLGTGENYTADSSQESLNLAAQTQVQKIEKTSQTSQQTQMLRDI 347

QY 348 SAFVKTPQWNKETENYSKGGEDHLOGGALLYVNDSTRTPWANSYRRLNRATATQGTI 407

Db 348 NSFVKTPQWNNSQTSDESAGEKDLQGGALLYSNSDKTAYANSYRLLNRPTSTQTK- 406

QY 408 DKSILDEQSDPNHMGDFPFLANDVDLSNPVVQAEQLNQIHYLMNMGSVIMGDKDANFDG 467

Db 407 -----PKYFEDNSSGCGYDFELLANDIDNSNPVVQAEQLNLHYLMNMGSVIVANDPEANFDG 461

QY 468 IRVDAVDNVDADMLQYNYFREYGVNKSSEANALAHISVLRAWSLNDNHNDKTDGAAL 527

Db 462 VRVDAVDNVDADLLQIASDYLKAHYGVDKSEKNAINHLISILEASDNDPNQYNKDTKGAQL 521

QY 528 AMENKQRLALLFSLAKPIK-----ERTPAVSPLYNNTFTNTQORDEKTDWINKDGSKAY 580

Db 522 PIDNKURLSLYALTRPLEKASKNKNEIRSGLEPVITNSLN----- 562
Qy 581 NEDGTQVIGKTYKNEKYGDASGVYFIRAHNNVDIIAELIKKEINPKSDGFTITDAE 640
Db 563 -----NRSREGKNSER---MANYIFIRAHSEVDQVIKIIKAQINPKTDGLTFLTDE 612
Qy 641 MKQAFELYNKMLSSDKKTYLNNIPAAAYVLMQMETITRVYVYGDLYTDGHHYMETKSPY 700
Db 613 LKQAFKTYNEDMRQAKKTYQSNIPATAYALMLSNKDSITRLXYGDMYSDDGQYMATKSPY 672
Qy 701 YDTIVNLMKSRKIKYVSGQARQSVLPTDCKMNSDVELYRTNEVYTSVRYGKDIMTAND 760
Db 673 YDAIDTLKARIKYAAGQDMKITVY--EGDKSHMD---WDYTGVLTSVRYGTCANEATD 727
Qy 761 TEGSKYSRTSGQVTLVANPEKMLDOSAKLNVEMGKIHANQYKRALIVGTADGINKFTSD 820
Db 728 -QSEATKTQGMAVITSNPESLKLNDQKVIYVNMGAHKKQERYPLLLITKDGLTSTYS- 785
Qy 821 ADAIAAGYKBTDSNGVLTFGANDIKGYETFDMSGFVAVVMPVPGASDNQDIRVAPSTEAK 880
Db 786 -DAAKSLYRKTNDKGELVFDASDIQGYLNPQVSGVLAVWVPVPGASDNQDVRVAASNKAN 844
Qy 881 KEGELTLKATEYDSQLIYEGFSPNQF-IPDGDSPSVYTRKIAENVDLPKSHGVTSEFEM 939
Db 845 ATGQV-YESSALDSQLIYEGFSPNQFVTKDSD---YTNKKIAQNVQLFKSWGVTSEFEM 900
Qy 940 APOFVSADGDTFLDSVTONGYAFADRYDLAMSKNNKYGSKEDLRDLAKLHAKAGIOAIAD 999
Db 901 APOYVSESDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSKQDDMINAVKALHKSIGIOVIA 960
Qy 1000 WPDQIYQLPGKEVVTATRTDAGRKIADAIIDHLSLYVANSKSSGKDYQAKYGGEEFLAEL 1059
Db 961 WPDQIYVLPKEVVTATRVNDYGEYRKDSEIKNTLYAANTKSNKGKDYQAKYGGAFISEL 1020
Qy 1060 KAKYPEFKNNMTSTGPKIDDSVKLKQWKAAYFNGTNVLERGVYVLSDEATKGYTTVTK 1119
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNILGRGVYVLDKONASDKYFELKG 1080
Qy 1120 EGNFIPQLTGKEKVTGTFSSDGKGYTGTSQTQAKSAFV-FTNGNTYVFDARGHMTN 1178
Db 1081 NQTYLPQMTNKE-ASTGFVNDGNGMTFTYSTGYQAKNSFVQDAKGNWYTFDNGNHWYV 1139
Qy 1179 SEYSPNGKDYVFLPENGIMLSNAFYIDANGNTYLYNSKGQYKGYTKFDVSETDKDGKE 1238
Db 1140 LQ-QLNG-EVQYFSLNGVQLRESFLENADGSKNYFGLGNRYNSGYYSFDNDS----- 1190
Qy 1239 SKVKFRYFTNEGVMKGVTVIDGFTQYFGBDGFQAKDLVT-FKGKTYVFDHAHTGNGIK 1297
Db 1191 ----KWRYFDASGVMAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDDGSGNMAV 1246
Qy 1298 DTWRN-INGKWWYFDVAGVAAGTGAOVINGOKLYFNEGDSQVKGVVKNADGTYSKYKEGF 1356
Db 1247 NFAUNDKNGDWYLYNSDGLALVGQTINGKTYVFGQDGKQIKGIITD-NGKLKYFLANS 1305
Qy 1357 GELVTNEFFTNGNWWYVYAGANGKTVTGAQVINGOHLNFNADGSQVKGVVKNADGTYSK 1416
Db 1306 GELARNIFATDSQNNWYVYFGSDGVAVTGSQTAGKLYFASDGKQVKGFSV-TYNGKVHY 1364
Qy 1417 YNASTGERUTNEFFTNGNWWYVYVYAGANGKTVTGAQVINGOHLNFNADGTYSKYKEGF 1476
Db 1365 YHADSGLQVNRPEADKNGWYVYLDNGBALTGSRINQVRVFTTREGKQVKGKQD-VAYDE 1423
Qy 1477 GRISY 1482
Db 1424 RLLVY 1429

RESULT 6
AAU98045
ID AAU98045 standard; protein; 1430 AA.
XX
AC AAU98045;

XX 27-AUG-2002 (first entry)
XX S. mutans glucosyltransferase GTFD mutant N471D/T589E.
DE Glucosyltransferase; GTFD; transgenic plant; paper sizing;
XX coating composition; glucan; starch; latex; thermoplastic molecule;
KW anyloplast; vacuole; paper manufacture; mutant; mutein.
XX Streptococcus mutans.
OS Synthetic.
OS
XX Key Location/Qualifiers
FH Misc-difference 471 /note= "Wild-type Asn substituted by Asp"
FT Misc-difference 589 /note= "Wild-type Thr substituted by Glu"
FT
XX US2002031826-A1.
XX 14-MAR-2002.
XX 19-DEC-2000; 2000US-00740274.
XX 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH/) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX Claim 36; Page: 44pp; English.
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which

CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36
XX
SQ Sequence 1430 AA;

Query Match 44.3%; Score 3510.5; DB 5; Length 1430;
Best Local Similarity 48.1%; Pred. No. 1.2e-179; Indels 101; Gaps 32;
Matches 725; Conservative 247; Mismatches 433;

QY 1 MENKHYKLVKVKQWVTVIAVASVALATVGLSVTTSSVSAD---ETQDKVTQ-SNSG 56
DB 1 METKRYKVKHVKHVVAVAS-GLITL--GTTILGSSVSAETEQOTSQDKVVTQKSEDD 57
QY 57 TTASLVTSPEATKEADKRTNTEADVLTPAKETNAVERATTTNQTATBAAT-ATTADV 115
DB 58 KAASESSQTDAPKTKQAQTEQTAQ-----SQANVADTSTSTKETPSONITTOANSDDK 112
QY 116 AVAAVPNKEAVTTDAPAVTTEKAEQAPATVKAEEVNVTEVKA-----PEALAKDSEVEA 169
DB 113 TVTNTKSEEAQTSSE-----RTQSEEAQTTASSQAL--TQAKAELTKQRTQAQENKNPVD 167
QY 170 ALSLNKINIKIDGKYVYVNDGSHKENFALTIVAGQLYFGKD--GALTSSSTYSFTPTGTNI 228
DB 168 LAAPNVKIDGKYVYVNDGSHKENFALTIVAGQLYFGKD--GALTSSSTYSFTPTGTNI 227
QY 229 VGFSSINNRAYSSEASFLIDGYLTADSWYRPAIIKDGVTWQASTABDFPLLMAMP 288
DB 228 NNDYTPHNIQVNPENTSETIDNYVTADSWYRPAIIKDGVTWQASTABDFPLLMAMP 287
QY 289 NVDTQVNYLYNYSKV--FNLDKAYSSTDQKQETLVKAAKDIQIKIEQIKQAEKSTOWLRETI 347
DB 288 DKQQTAYLYNMQOGLGTGENYATDSSQESLNLAQTVQVKIETKISQTOQTOWLRDII 347
QY 348 SAFVKTQPOWNETENYKSGGDEHLOGGALLYVNDRTPWANSYRRLNRTATNQTGTI 407
DB 348 NSFVKTQPNWNSQTESDTSAGEKDLHOGGALLYVNDRTPWANSYRRLNRTATNQTGTI 406
QY 408 DKSILDEQSDPNHMGDFLLANDVLSNPVQAEOLNOIHYLMNWSIVMGDKDANFDG 467
DB 407 -----PKYFEDNSSGGYDFLLANDVLSNPVQAEOLNOIHYLMNWSIVMGDKDANFDG 461
QY 468 IRVDVNDVADMLQLYTVFREYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAAL 527
DB 462 VRVDVNDVADMLQLYTVFREYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAAL 521
QY 528 AMENKQRLALLFLSLAKPIK-----ERTPAVSPLYNNTFTTQRTDEKTDWINKDGSKAY 580
DB 522 PIDNKLRLSLLYALTFPLEKADSKNEIRSGLEPVITNSLN-----562
QY 581 NEDGTQVQSTIKYNEKYGDASGVYFIRAHNNVDIIAEIIKKEINPKSGDFTITDAE 640
DB 563 -----NRSAGEKNSER-----MANYIFIRAHNSEVQEVIAKIKAIQNEPKTDGLFTLDE 612
QY 641 MKQAEPIYKNDMLSSDKKYTLNPIPAAYAVMLQNMETITFRVYGDLYTDGHHMETKSPY 700
DB 613 LKQAFKIYNEDMRQAKKYTKQNIPTAYALMISNKDSITRLYGDYSDDGQYMATKSPY 672
QY 701 YDTIVNLMKSRIKYSGGQARSLPMTDQKMDNSDVELYRTNEVVTYSVRYGKIDIMTAND 760
DB 673 YDAIDTLKARIKYAGGDMKITTV--EGDKSHMD---WDYTGVLTSVRYGTGANEARD 727
QY 761 TSGSKYSRTSGQVTLVANNPKNLNDQSAKLNVEMGKIHANQKYRALIVGTAGIKNFTSD 820
DB 728 -QGEATKTKQMAVITSNPNPSLKLNDQKIVVMGAAHKNQEVYRPLLLATTKDGLTSYTS- 785
QY 821 ADAIAAGYVYKETSNGVLTFGANDIKGYETFDMSGFPVAVVVPVGSADNQDIRVAPSTEAK 880

DB 796 -DAAAKSYLRKINDKGLVDFDASDIQGYLNPVDSGLVAVVPGASDQDVRVAASNKAN 844
QY 881 KEGELTLKATEAYDSQLIYEGFSNFQT--IPDGSDPVSVTNRKIAENVDLFSKSGVTSPEM 939
DB 845 ATGQV-YESSSALDSQLIYEGFSNFQDFVTKDSD--YTNKKIAQNVLFSKSGVTSPEM 900
QY 940 APQFVSADGTFPLDSVIONGVAFADRYDLAMSKNNKYSKEDRLDALKALHAGIQAID 999
DB 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYSQDDMINAVKALHKSIGIQAID 960
QY 1000 WYPDOIYOLPGKEVVTATRTDAGRKIADAIIDHSLYVANSKSGDYQAKYGGCFEASL 1059
DB 961 WYPDOIYOLPGKEVVTATRTDAGRKIADAIIDHSLYVANSKSGDYQAKYGGCFEASL 1020
QY 1060 KAKYEMFKVNMISTGKPIDDSVKLQWKAEBYFNGTNNLVRGVYVLSDEATGKFTVTK 1119
DB 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNNLVRGVYVLSDEATGKFTVTK 1080
QY 1120 EGNFTPLQLTGKEKVIITGSSDGKGIITFGTSGTOAKSAFV--TFNGNTYTFDARGHMTN 1178
DB 1081 NQYILPKQMTNKE--ASTGFVNDGNGMTFYSTSGYQAKNSFVQDAGKNWYTFDNGHMYG 1139
QY 1179 SEYSPNGKDVYRFLPENGIMLSNAPYIDANGNTYLYNSKGOMYKGYTKFDVSETDKDGE 1238
DB 1140 LQ-QLNG-EVQYFLSNGVQLRESFLENADGSKYFCHLGNRYNSGYSFNDNS----- 1190
QY 1239 SKVKRYFTNEGVMNAKVTVIDGTFQYFGEDEGQAKDLVT--FKGKTYTFDAHTNGIK 1297
DB 1191 ---KWRYFDASGVMAVGLKTINGNTQYFDQDGYQVKGAWITGSDGKRYFDDGSGNMAV 1246
QY 1298 DTRWN--INGKWTYFDANGVAATGAQVINGOKLYFNEDESQVKGVVKNADGTYSKYKEGF 1356
DB 1247 NRFANDKQGDWYLYNSDGIALVGQTINGKTYTFQDQKQIKGKIITD--NGKLKYFLANS 1305
QY 1357 GELVNEFEFTDGNVWYVAGKVTGTAQVINGOHLYFNADGSOVKGGVKNADGTYSK 1416
DB 1306 GELARNIFATDSQNNWYTFSGDVAVTGQTTAGKLLFPASDGKQVKSFV--TYNGKVHY 1364
QY 1417 YNASTGERLTNEFTTGDNNWYIIGANGKSVTGEVIGKIDDTYFFAKDGKQVKTVSAGN 1476
DB 1365 YHADSGELQVNRFEADKQGNWYLYNSDGEALTGSRINDQRVFTREGKQVKG--VAYDE 1423
QY 1477 GRISY 1482
DB 1424 RLLVY 1429
RESULT 7
AAU98041
ID AAU98041 standard; protein; 1430 AA.
XX
AC AAU98041;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFD mutant TS89D.
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutin.
XX Streptococcus mutans.
OS Synthetic.
XX Key Location/Qualifiers
FT Misc-difference 589
FT /note= "wild-type Thr substituted by Asp"
XX US2002031826-A1.
XX 14-MAR-2002.
XX


```

QY 1179 SEYSPNGKDVYRFLPNGIMLSNAPYIDANGNTYLYNSKQMGYKGGYTKFDVSETDKDGE 1238
Db 1140 LQ-QLNG-EVQYFLSNGVQLRESFLENADGSKNYFGHLGNRYSNGYSPDNDS----- 1190
QY 1239 SKVWFRYFTNBSGVMAGVTVTDGFTQYFGEDEGFOAKDKLVT-FKGKTYFFDAHTGNGIK 1297
Db 1191 ----KWRYFDASGMAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKYFDDGSGNMAV 1246
QY 1298 DTWRN-INGKYYFFDANGVAATGAQVINGOKLYFNSDGSQVKGVVKNADGTSYK 1356
Db 1247 NRPANDKNGDWYLYNSDGLGVQTINGKTYFFQDQKIGKIITD-NGKLYFLANS 1305
QY 1357 GELVTNEFTTGNVWYAGANGKTVTGAOVINGQHLNYFNADGSQVKGVVKNADGTSYK 1416
Db 1306 GELARNIFATDSQNNWYFSGDVAVTGSGTTAGKLYFASDGKQVKSFV-TYNGKVHY 1364
QY 1417 YNASTGERLTNEFTTGDNNWYIYGANGKSVTGEVKIGDDTYFFAKDGKQVKGQTVSAGN 1476
Db 1365 YHADSGELQVNRFEADKQGNWYILDSNGEALTGSRINDQRVFFTRGKQVKGQD-VAYDE 1423
QY 1477 GRISYY 1482
Db 1424 RLLVY 1429

RESULT 8
ID AAU98043 standard; protein; 1430 AA.
XX
AC AAU98043;
XX
DT 27-AUG-2002 (first entry)
XX
S. mutans glucosyltransferase GTFD mutant N471D.
XX
KW Glucosyltransferase; GTFD; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
XX
SY Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 471 /note= "Wild-type Asn substituted by Asp"
XX
XX US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
PA (NICH/) NICHOLS S E.
XX
PI Nichols SE;
XX
XX WPI; 2002-414332/44.
XX
PT Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
XX Claim 36; Page; 44pp; English.
XX
XX

```

```

CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from 1448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFD mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFD sequence appearing as AAU98029 and the information in
CC claim 36
XX
XX Sequence 1430 AA;
XX
Query Match 44.2%; Score 3507.5; DB 5; Length 1430;
Best Local Similarity 48.1%; Pred. No. 1.7e-179;
Matches 725; Conservative 246; Mismatches 434; Indels 101; Gaps 32;
QY 1 MENKIYKHLKVKQWVTIAVASVALATVGLGSLVTTSSVSAD---ETQDKTVTQ-SNSG 56
Db 1 METKRYKWHKVKKHMTVAVAS-GLITL--GTTTLGSSVSAETEQTQSDKVVTKQSEDD 57
QY 57 TTASLVTSPKATKADKRTNTEKADVLTPAKETNAVEATTTNTQTAEAAAT-ATTADV 115
Db 58 KAAESSQTDAPKTKQAQTEQTAQ-----SQANVADTSTSTITKTPSQNITTAQNSDDK 112
QY 116 AVAAVPNKEAVVTTDAPAVTTTEKAEQPAATVAAEVNTEVKA-----PEAALKDSEVEA 169
Db 113 TVTNTKSEEAQTSEE-----RTKQSEEAQTASSQAL-TOAKAELTKQRTAQENKNPVD 167
QY 170 ALSLKNIKNIDGKYIYVNEDEGSHKENFAITVNGQLLYFGKD-GALTSSTSYSTFTPGTTNI 228
Db 168 LAAPVNVKQIDGKYIYVNEDEGSHKENFAITVNGQLLYFGKD-GALTSSTSYSTFTPGTTNI 227
QY 229 VDGFSINNRAYDSSEASFELIDGYLTADSWYTPASIIKGVQVWQASTADDFRPLLMANWP 288
Db 228 NNDYTPHNQIVNFENTSLTETIDNYVTADSWYTPASIIKGVQVWQASTADDFRPLLMANWP 287
QY 289 NYDTQVNYLYNYSKV-FNLDAKYSTDQKQETLKAQKIQIKIEOKIQAEKSTQWLRETI 347
Db 288 DKQTQIAYLYNYSKV-FNLDAKYSTDQKQETLKAQKIQIKIEOKIQAEKSTQWLRETI 347
QY 348 SAFVKTPQPMNKETENYKGGGDEHLQGGALLYVNDSTRTPWANSYDRLNRRATNATGTGT 407
Db 348 NSFVKTPQPMNQSQTESDTSAGEKDLQGGALLYVNDSTRTPWANSYDRLNRRATNATGTGT 406

```


Db 901 APQVSSDGSFLDSIIQNGYAEDRYDLAMSKNNKYGSGQDMINAVKALHKSQIQUIAD 960
Qy 1000 WVPDQIYQLPGKEVWVATRTDGRKTIADAIIDHSLYVANSKSGDYQAKYGEFLAEI 1059
Db 961 WVPDQIYQLPGKEVWVATRVNDYGEYKDSKSEIKNTLYAANTKNGKDYQAKYGAFLSEL 1020
Qy 1060 KAKYPEMKVNMISTGKPIDDSVKLQWKALYFNGTNVLRGYYVLSDATQKTYFTVTK 1119
Db 1021 AAKYPSIFNRTQISNGKKIDPSKITAUKAKYFNGTNILRGVGYVLKONASDKYFELKG 1080
Qy 1120 EGNFIPQLTGKKEKVTGFSDDGKITFGTSGTOAKSAFV-TFNGNTYYFDARGHMTN 1178
Db 1081 NQTYLPKQMTNKE-ASTGFVNDGNGHTFYSTGYQAKNSFVQDAKGNWYTFDNGHMYG 1139
Qy 1179 SEYSPNGKDYVRLPNGIMLSNAPYIDANGNTLYLNSKGMYKGYTKFDVSETDKDGE 1238
Db 1140 LQ-QLNG-EVOYFSLNGVQLRESFLENADGSKNYFGLGNRYNSGYSPDNDS----- 1190
Qy 1239 SKVVKERYFTNEGMAKGVTVIDFTQYFGEDGFOAKDKLVT-FKGKTYTFDHTNGIK 1297
Db 1191 ----KWRYFDASGMAVGLKTINGNTQYFDQDGYQVKGAWITGSGKRYFDGSGNMAV 1246
Qy 1298 DTWRN-INGKWWYFDANGVAATGAOVINGOKLYFNEGDSQVKGGVKNADGTYSYKEGF 1356
Db 1247 NRPANDKGDWYIYNSDGLVGVQTINGKTYFYFGDQKQIKIITD-NGKLKYFLANS 1305
Qy 1357 GELVTNEFFTTDGNVWYAGANGKTVTGAQVINGOHLNFADGSGVKGGVKNADGTYSK 1416
Db 1306 GELARNIFATDSQNNWYFGSDGVAVTGSGTIAGKLYFASDGKQVKGFSV-TYNGKVHY 1364
Qy 1417 YNASTGERLTNEFTTGDNNWYIYGANGSVTGEVKIGDDTYFFAKDGKQVKGQTSYAGN 1476
Db 1365 YHADSGELQVNRFEADKGNWYIYLDNGEALTGSRINDQRVFTREGKQVKG-D-VAYDE 1423
Qy 1477 GRISY 1482
Db 1424 RRLVY 1429

RESULT 12
ADX37279
ID ADX37279 standard; protein; 1430 AA.
XX
AC ADX37279;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #8.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
XX WPI; 2005-151644/16.

XX
PT New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
PS Claim 7; SEQ ID NO 36; 73pp; English.
XX
CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
SQ Sequence 1430 AA;

Query Match 44.2%; Score 3502.5; DB 9; Length 1430;
Best Local Similarity 48.1%; Pred. No. 3.1e-179;
Matches 724; Conservative 247; Mismatches 434; Indels 101; Gaps 32;
Qy 1 MENKIHVHLKHKVKKWVTIIVASVALATVGLSVTTSSVSAD---ETODKTYTO-SNSG 56
Db 1 METKRRYKHKVKKWVTIVAVAS-GLITL--GTTILGSSVSVAETEQTSDKVVTKSEDD 57
Qy 57 TTASLVTSPEATKEADKRTNTKEADVLTPAKETNAVETATTTNTQTAAEAAT-ATTADV 115
Db 58 KAASESSQTDAPKTKQAQTEQTAQ-----SQANVADTSTSIKETPSQNIITQANSDDK 112
Qy 116 AVAAVNPKEAVVTDAPAVTTBKAEPATVKAENVNTEVKA-----PEAAUKDSEVEA 169
Db 113 TVTNTKSEEAQTSSE-----RTKQSEEAQTASSQAL-TQAKAELTKQRTAAQENKPNVD 167
Qy 170 ALSLKNIKNDGKYVYNEDGSHKENFALTIVGQLLYFGKD-GALTSSTSYSTPGTTNI 228
Db 168 LAAIPNVKOIDGKYIYIGSDGPKNFALTIVNKKLVYFDKNTGALTDTSTGYQKQGLTKL 227
Qy 229 VDGFSINNRAYDSSEASFELIDGYLTADSWYPASIIKDGVTWQASTADREFLLMAWVP 288
Db 228 NNDYTPHNQIVNFENTSLETIDNYVTADSWYRPKDILKNGKTWTASSEDLRLLMSWVP 287
Qy 289 NVDTQVNLNYSKV-FNLDAKYSTDKQETHKVAAKDIOIKIEQIKIAEKSTQWLRETI 347
Db 288 DKQTQIAYLNNVNNQGLGTGENYTDSSQESLAAQTQVVKIETKISQTOQTQWLRLDII 347
Qy 348 SAFVKTQPOWNETENYSKGGEDHLQGGALLVYVNDSTRTPWANSVRRNRTATNQTGTI 407
Db 348 NSFVKTQPNWNSQTSBDSAGBKDHLQGGALLYSNDSKTYANSVYRLNRTPTSGTK- 406
Qy 408 DKSILDEQSDPNHMGDFDPLANDVDLSNPVVOAEQNLQIHVLMNWSIVMGDKDANFDG 467
Db 407 -----PKYFEDNSSGGYDFLLANDIDNSNPVVOAEQLWLHLYMNGSVIVANDPEANFDG 461
Qy 468 IRVDADVNDADMLQLYNTYFREYGVNKSANALAHISVLEAWSLNDNHNDKTDGAAAL 527
Db 462 VRVDADVNVNADLLQIASDYLKAHYGVDSKSEKNAIHLISLEAWSNDPQYNKDKTKGAOL 521
Qy 528 AMENKQRLALFSLAKPIK-----ERTPAVSPLYNNTFNTTQORDEKTDWINKDSKAY 580
Db 522 PIDNKLRLSLLYALTRPLEKODASNKNETRSGLPEVITNSLN----- 562
Qy 581 NEDGTGVKOSTICKYNEKYGDAAGNVFFIRAHNNVQDIIAEIIEKINPKSGDFTITDAE 640
Db 563 -----NRSAEKNSEK-----MANYIFIRAHNSEVQTVIAKIIKAQINPKTDGLTFLDE 612
Qy 641 MKQAFEINYKMDLSSDKYTLNIIIPAAVAVMLQNNMETITRVYGYDLYTDGHHMETKSPY 700
Db 613 LKQAFKIVNEDMRQAKKYTQSNIPATVALMLSNKDSITRLVYGDWYSDGQVMAKSPY 672
Qy 701 YDTIVNLMSRKIKYVSGGAQORSYWLPTDGKMDNSDVELYRTNEVYTSRYGKIDINTAND 760
Db 673 YDAIDTLKARIKYAAGGQDMKITTV--EGDKSHMD----WDYTGVLTSRYGTGANEATD 727

Qy	1244	FRYFTNEGVMAGVTVIDGFTQYFGEDGFOAKDK-LVTFPKGTYFYFDAHTGNGIKDWT-R	1301
Db	1158	WRIFY-KDGNNAVGLTITVDGNQYDFDKQVQAKDKIIVTRDGKRVFDQHNGNAVNTFIA	1216
Qy	1302	NINGKWIYFDANGVAATGAQVINGOKLYFNEGDSQVKGSGVVKVKNADGTYYSKYKEGFGELVT	1361
Db	1217	DKTGHWYLYLKGQGA VGTGAQTGVKQKLYPEANGQQVKGDFVTSDEGKLYFYDVDSDMMWT	1276
Qy	1362	NEPFTTDDGNVWYIYAGANGKTVTGAOVINGOHLIYENADGSOVKGVVKNADGTYYSKYKNAST	1421
Db	1277	DTFIEDKANWVYLGKDGGAATGAQTIRGQKLYFKANGQQVKGDIVKGTGDKIRIYDASK	1336
Qy	1422	GERLTNE-----PFT	1431
Db	1337	GEQVFNKTKAADGKTYIENDGVA VDPSPVVKGTFFKQASGALRFPYNLKGQLVTGSGWYE	1396
Qy	1432	TGDNNNWYIYGANGKSVTGRVKTKGDDTYFPAKDGKOVKGQTVYSAGNGRISYYVYDGSQKRAV	1491
Db	1397	TANHDMWYI-QSGKALTGQBTINGOHLIYFKPDGHQVKGLVVTGTGDKRVYYDANGSDQAF	1455
Qy	1492	STWIEIQPGVYVYFDKNGLA	1511
Db	1456	NKSVTVNGKTY-YFGNDGTA	1474

Search completed: February 11, 2006, 19:15:26
Job time : 159.097 secs

Qy	172	SLKWIKNIDGKYYIYNVEDGSHKENFALTUNGQLLYFGKOGALTSSSSYFTPTGTTNIDVG	231
Db	150	KMSNVKQVDGKYYIYDQDGNVKNFVSVGDKIYYFDETGAYKDTSKVDADKSSVASQN	209
Qy	232	---FSINNRAYDSSEASFELIDGLYTADSWRSPASIIKOGVTWOASTAEDFRPLLMAWP	288
Db	210	ATTPAANNRAYTSACKFEADVNYLTADSWRPKSILKOGKTWESGKODFRPLLMAWP	269
Qy	289	NVDTQVNYLNYMSKVFNLDAKYSSTDQKQETLKVAAKDIQIKIEQIOAEKSTOWLRETIS	348
Db	270	DTETKRNYYNVMKVVGIDKTYTAETSAQDITAAAEVLQARIEQKITSENNKTWLEAIS	329
Qy	349	AFVKTQOWNKETEN--YSKGGEDHLOGGALLYVNDY--RTPWANSYVRRLLNRATNQGT	406
Db	330	AFVKTOPMNGESEKPY-----DDHLQNGALLFDNQDLDLTPDQSNRYLLNRPTNQTGS	384
Qy	407	IDKSILDEQSDPNHMGSGDFLLANDVDLSPNVQAEQLNOIHYLMWGSIVMGDKDANFD	466
Db	385	LDGRFTYTPNDP--LGGYDFLLANDVDSNPVQAEQLNWLHLNLLFGSIYANDADANFD	442
Qy	467	GIRVDAVDNDVADMQLYTYNNFREYYGVNKSEANALAHISVLEAWSLNDNHYNDKTDGAA	526
Db	443	SIRVDAEDNDVADQLOISSDYLKAAIGIDKNKNANNHVSIVRAWSNDPTPYLHDDGDNL	502
Qy	527	LAMENKORLALLPSLAKPIKERTPAVSPLYNNFTNTTQRDEKTDWINKOGSKAYNEDGTV	586
Db	503	MNMNDKPRLSMLSLAKPTDVR--GLNPLIHNLSLVREVDDR-----EVETV	548
Qy	587	KQSTIGKYNKYGDASNYVPIRAHDNNVODITAEIIEKKEINPKSGDFTITDAEMQOAPE	646
Db	549	P-----SYSFARAHSEVQDITRDIKAEINPNSFGYSTQEIIOAEK	592
Qy	647	IYNKDMLSDBKTYLNNIPAAAYAMLQNMTETIRVYYGDLYTDDGHVMEFKSPYDTIVN	706
Db	593	IYNEDLKSDKKYTHYNVPLSYLLLTNKGSIIPRVYGDNFTDQGYMANKTVNYDAIES	652
Qy	707	LMKSRIKYVSGGOQRSWMLPTDGMDSNDVELYRTNEVYTSVRYGKDIIMTANDTBGSKY	766
Db	653	LLKARMKYVAGGQAMQNYI-----GNGEILTYSVRYKGALKQSD--KGDAT	697
Qy	767	SRTSGQVTLVANNPKMLDOSAKLVNEMGKIHANQKTRALI VGTADIGIKNFTSDADAIAA	826
Db	698	TRTSGYGVVNGNPNFSLDGKV--VALMNGAAHANQBYRALMVSTKDGVAITYADADASKA	756
Qy	827	GYVKETDSNGLVFGANDIKGYETFDMSGFVAVVVPVPGASDNODIRVAPSTEAKKSGELT	886
Db	757	GLVKRTDENGLYLFLNDDLKGANPNQVSGFLVWVPVGAADDQDIRVAASDSTASTDGK--S	815
Qy	887	LKATEAYDSQLIYEGFSNFOTIPDGDSPSYVYNNRKAENVDLPKSGVTSFEMAPQFVSA	946
Db	816	LHODAAMDVRMVEGFSNFOSP--ATKEBEYTVNVIANNVDKFSWGITDFFENAPQVSS	873
Qy	947	DDGTFLDSVTQNGYAFADRYDLAWSKNKYSKESKEDLRDALKALHKAGIOAIADWPQIY	1006
Db	874	TDGOFODSVIQNGYAFTRDYLGWSKANKYGTADQLVKAIKALHAKGLKVMADWPQOMY	933
Qy	1007	QLPGKEVWYATRTDAGRKRTADAILDHSLLVANSKSGKDYQAKYGEFELAEIKAKYPBM	1066
Db	934	TPPKQEVVYVTRTDKFGKPIAGSQIHNLSLVYTDPTKSGDDYQAKYGEAFDELKEKYPKL	993
Qy	1067	EKYNMLSTGKPIDDSVKLKQWKAEYFNGTNVLBERGVGYVLSDBATGKYFTVTKEGNFPL	1126
Db	994	FTKKQSTGQAIIDPSVKIKQWSAKYFNGSNILRGADYVLSQVSNKYFNVASDITFLPS	1053
Qy	1127	QLTGKEKVITGFSDDGKITY--FGTSGTOAKSAFVTFNGNTYYTFDARGHMVTNSEYSPNG	1185
Db	1054	SLLGK--VWESGIRYDCKGYIYNSSATGDQVKASFIETAGNLYYFGKDGYYVMTCAQ--TING	1111
Qy	1186	KDYTRFLPNCIMLSNAPYIDANGNTYLYNSKGQMY--KGYTYTFDVSDETDDOKESKVKV	1243
Db	1112	AN--YFLENGTALRNTIYTDQGNSHYVANDGRYENENGYQOF-----GND-----	1157

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 33.2355 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-40

Perfect score: 7928

Sequence: 1 MENKIHYKLHKVKKQWVTIA.....PGVYVYFDKNGLAYPRVLN 1518

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/1aa/5 COMB.pep.*

2: /cgn2_6/prodata/1/1aa/6 COMB.pep.*

3: /cgn2_6/prodata/1/1aa/H COMB.pep.*

4: /cgn2_6/prodata/1/1aa/PCTUS COMB.pep.*

5: /cgn2_6/prodata/1/1aa/RE COMB.pep.*

6: /cgn2_6/prodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3502.5	44.2	1430	2 US-09-008-172-2	Sequence 2, Appli
2	3502.5	44.2	1430	2 US-09-210-361-6	Sequence 6, Appli
3	3502.5	44.2	1430	2 US-09-740-274-6	Sequence 6, Appli
4	3196.5	40.3	1475	2 US-09-007-999-2	Sequence 2, Appli
5	3196.5	40.3	1475	2 US-09-210-361-2	Sequence 2, Appli
6	3196.5	40.3	1475	2 US-09-740-274-2	Sequence 2, Appli
7	3155	39.8	1577	1 US-08-793-824-2	Sequence 2, Appli
8	3070	38.7	1375	2 US-09-210-361-4	Sequence 4, Appli
9	3070	38.7	1375	2 US-09-740-274-4	Sequence 4, Appli
10	2327	29.4	2057	2 US-09-499-203-2	Sequence 2, Appli
11	2093	26.4	1278	2 US-09-604-957-3	Sequence 3, Appli
12	2093	26.4	1781	2 US-09-995-749A-2	Sequence 2, Appli
13	1584.5	20.0	545	2 US-09-604-957-4	Sequence 4, Appli
14	1560.5	19.7	545	2 US-09-995-749A-10	Sequence 10, Appl
15	1262.5	15.9	523	2 US-09-604-957-5	Sequence 5, Appli
16	1262	15.9	522	2 US-09-995-749A-11	Sequence 11, Appli
17	1220.5	15.4	535	2 US-09-604-957-7	Sequence 7, Appli
18	1220.5	15.4	535	2 US-09-995-749A-13	Sequence 13, Appl
19	1090	13.7	584	2 US-09-604-957-6	Sequence 6, Appli
20	1088	13.7	584	2 US-09-995-749A-12	Sequence 12, Appl
21	586.5	7.4	349	2 US-09-009-620-2	Sequence 2, Appli
22	523	6.6	2710	1 US-08-480-604A-6	Sequence 6, Appli
23	523	6.6	2710	1 US-08-405-496A-6	Sequence 6, Appli
24	523	6.6	2710	2 US-08-915-136-6	Sequence 6, Appli
25	523	6.6	2710	2 US-08-957-310-6	Sequence 6, Appli
26	523	6.6	2710	2 US-10-011-366-6	Sequence 6, Appli
27	523	6.6	2710	2 US-09-084-517-6	Sequence 6, Appli

28	447	5.6	811	1	US-08-480-604A-7	Sequence 7, Appli
29	447	5.6	811	1	US-08-405-496A-7	Sequence 7, Appli
30	447	5.6	811	2	US-08-915-136-7	Sequence 7, Appli
31	447	5.6	811	2	US-08-957-310-7	Sequence 7, Appli
32	447	5.6	811	2	US-10-011-366-7	Sequence 7, Appli
33	447	5.6	811	2	US-09-084-517-7	Sequence 7, Appli
34	447	5.6	812	1	US-08-480-604A-29	Sequence 29, Appl
35	447	5.6	812	2	US-08-915-136-29	Sequence 29, Appl
36	447	5.6	812	2	US-09-084-517-29	Sequence 29, Appl
37	436.5	5.5	866	2	US-09-545-773-2	Sequence 2, Appli
38	436.5	5.5	866	2	US-10-222-038-2	Sequence 2, Appli
39	379.5	4.8	1231	2	US-08-714-741-41	Sequence 41, Appl
40	366	4.6	2366	1	US-08-480-604A-10	Sequence 10, Appl
41	366	4.6	2366	1	US-08-405-496A-10	Sequence 10, Appl
42	366	4.6	2366	2	US-08-915-136-10	Sequence 10, Appl
43	366	4.6	2366	2	US-08-957-310-10	Sequence 10, Appl
44	366	4.6	2366	2	US-10-011-366-10	Sequence 10, Appl
45	366	4.6	2366	2	US-09-084-517-10	Sequence 10, Appl

RESULT 1
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match	44.2%	Score	3502.5	DB 2;	Length	1430;			
Best Local Similarity	48.1%	Pred. No.	4.7e-222;	Mismatches	434;	Indels	101;	Gaps	32;
Matches	724;	Conservative	247;						

Qy	1	MENKIHYKLHKVKKQWVTIAVASVALATVLGGLSVTTSSVSAD---ETQDKTVTQ-SNSG	56
Db	1	METKRRYKMHKVKHWHVTVAVAS-GLITL--GTTTLGSSVSVAETEQTSDKVVTKSDD	57
Qy	57	TTASLVTSPKATKADKRTNKEADVLPKAKETNAVEATTNTQATABAAT-ATTADV	115
Db	58	KAASESSQTDAPKTKQAQTEQTAQ-----SQANVADTSTSTKETPSQNITTQANSDDK	112
Qy	116	AAVAPNKEAVVTTDAPAVTTKEAEQPAVKAEVNVTEVKA-----PEALKDSEVEA	169
Db	113	TVNTKSEEAQTSSE-----RTKQSEEAQTASSQAL--TQAKAELTKQRTAQAKNPNVD	167
Qy	170	ALSLKNIKNIDGKYIYVNEGSHKENFAITVNGQLLYFGKD--GALTSSTSYSTFTPTTNI	228
Db	168	LAAPNVAQIDGKYIYIGSDGQPKNFALTNNKVLKFDKNTGALTDTISQYQKQGLTKL	227
Qy	229	VDGFSINNRAVDSSEASPELIDGYLTADSWRYPASIIKGVVTWQASTABDFRPLLMWMP	288
Db	228	NNDYTPHNQIVNFENTSLETIDNYVTADSWRYPKDKILKNGKWTWASSSEDLRPLMSWP	287
Qy	289	NVDVTQVNLNYSKV-FNLDAKYSSTDQKFTLKVAAKDIQIKIEOKIOAEKSTOWLRETI	347
Db	288	DKQTOIAYLNNYNQOGLGTGENYTADSSQESLNLAQTQVQVKIETKISQTOQTOWLRDI	347
Qy	348	SAFVKTOPWNKETENYSKGGEDHLOGGALLVYVNSRTPWANSYRRLNRTATNTQGTI	407

ALIGNMENTS


```
QY 528 AMENKORLALLSLAKPIK-----ERTPAVSPLYNNFTNTQDEKTDWINKGSKAY 580
Db 522 PIDNKLRLSLYALTRPLEKDRASKNNEIRSGLEPVITNSLN-----562
QY 581 NEDGTQKQSTIGIKYNEKYGDASGNVYFIRAHNNVQDIIAEIIEKEINPKSDGFTITDAE 640
Db 563 -----NRSAEGKNSER-----MANYIFIRAHSEVQTVIAKIIKAQINPKTDGLFTILDE 612
QY 641 MKQAFIYNKMLSSDKKTYTLANNIPAAVAMLQNMETITRVYVYDGLYDGDGHMETKSPY 700
Db 613 LKQAFKIYNEDMRQAKKYTQSNIPYALMLSNKDSITRLYGDWYSDGQYMATKSPY 672
QY 701 YDITVNLMSRLKIYVSGGQAQSYWLPDQKMDNSDVELYRTNEVYTSVRYKDIWAND 760
Db 673 YDAIDTLKARIKYAAGGQDMKITVY--EGDKSHMD---WDYTGVLTSVRYGTGANEATD 727
QY 761 TEGSKYSRTSGQVTLVANNPKLNLQDSAKLNYEMGKIHANQYKRALIVGTADGIKNFTSD 820
Db 728 -QGSEATKTOGMNAVITSNPSLKLNQDKVIVNMGNAHKQYRPLLLTKDGLTSYTS- 785
QY 821 ADAIAAGYVYKETSNGVLTFGANDIKGYETFDMSGFVAVVMPVGASDNQDIRVAPSTEAK 880
Db 786 -DAAAKSLYRKTNKGELVFDASDIQGYLNPQVSGYLAVMVPVGASDNQDVRVAASNKAN 844
QY 881 KEGELTLKATEAYDSOLIIYEGFSNQT--IPDGSDDPSVYTNRKIAENVDLFKSWGVTSPM 939
Db 845 ATGQV--YESSSALDSOLIIYEGFSNQQDFVTKDS--YTNKKAQNVQLFKSWGVTSPM 900
QY 940 APQVYSADGTGLDSDVIQNGYAFADRYDLAMSKNNKYGSKEDRLDALKALHKAGTQAIAD 999
Db 901 APQVYSSEDSGLSDSIQNGYAFEDRYDLAMSKNNKYGSQQDMINAVKALHSGGLQVIAD 960
QY 1000 WVPDQIYQLPGKEWVATRTDAGAKRIADAIIDHSLYVANSKSGKDYQAKYGGSEFLAEL 1059
Db 961 WVPDQIYNLPGKEWVATRVNDYGEVRKDSIEIKNTLYAANTKSNKDYQAKYGGAFLEL 1020
QY 1060 KAKYPMPKVMNISTGKPTDDSDVKLQWKAEBFNGTNVLERGVYVLSDEATGKFTVTK 1119
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTNIILGRGVYVLDKNASDKYFELKG 1080
QY 1120 EGNFPLQLTGKEKVTGFSFGKGTTFGTSTQAKSAFV--TFNGNTYFDFARGHMTN 1178
Db 1081 NOTYLPKQWNTKE--ASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYTFDNGHMYG 1139
QY 1179 SBYSNPGKDYVYRPLNPGIMLSNAFYIDANGNTYLYNSKQMYKGGYKTFDVSETDKGKE 1238
Db 1140 LQ--QLNG--EVQVFLSNGVQLRESFLENADGSKNYFGHLGNRYSNGYSPDNDS-----1190
QY 1239 SKVVKFRYFTNKGVMNAKGVTVIDGFTQYFEGEDGPOAKDKLVT--FKGKTYVFDAGTNGIK 1297
Db 1191 ----KWRYFPDASGVMAVGUKTNGNTQYFDQDGQVQKGAWITGSDGKGRYFDGSGNMAV 1246
QY 1298 DTWRN--INGKYYVDFDANGVAATGAOVINGOKLYFNEBDSQVKGGVVKNADGTVSKYKEGF 1356
Db 1247 NRPANDKXGDWYLYNSDGLALGVQTINGKTYTFQDQGHQIKGKIITD--NGKLYFLANS 1305
QY 1357 GELVTFNEFTDGNVWYVYAGANGKTVTGAQVINGQHLRYFNADGSOVKGGVKNADGTSK 1416
Db 1306 GELARNIFATDSQNNWYVYFGSDGVA VTGSQTIAGKLYFASDGKQVKGSGFV--TYNGKVHY 1364
QY 1417 YNASGERLTNEBFTTGDNNWYVYIGANGKSVTGEKIGDDTYFFAKDGKQVKGQTVSAGN 1476
Db 1365 YHADSGELQVNRFEADKQGNWYLYDNGSEALTGSGRINQORVFFFTREGQVKGVD--VAYDE 1423
QY 1477 GRISY 1482
Db 1424 RALLVY 1429
```

RESULT 4

US-09-007-999-2

; Sequence 2, Application US/09007999

; Patent No. 6087559

```
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-007-999-2
```

```
Query Match 40.3%; Score 3196.5; DB 2; Length 1475;
Best Local Similarity 45.2%; Pred. No. 7.5e-202; Indels 139; Gaps 31;
Matches 695; Conservative 233; Mismatch 472;

QY 1 MENKHYKLHKYKQWVTIAVAS--VALATVLGGLSVTTSSVSADBTQDKTVTQNSGTT 58
Db 1 MDKVKYKLRKYKRWVTVSVASAVMTLTTLSSGL-----VKADSNESKQISNDSNTS 54
QY 59 ASLVTSPKATKADKRTNKEADVLTTPAKETNAVETATTTNTQATAAATATTADVAVA 118
Db 55 -----VVTANEESNVITEATSQBAASQTNHVTVTSSSSTS 91
QY 119 AVPNKEAV---VTTDAPAVTTEKAEQEPATV---KAEVVNTEVKAPEA---ALKDS--- 165
Db 92 VVNPKEVSNPYTVGETASNGEKQNTTVDKTSEAAANNISKQTEADTVIDDSNAA 151
QY 166 EYEAALSUKNIKIDGKYVYVNEDESHKENFAITVNGQLLYFGKQDGLTSSSTYSPG- 224
Db 152 NIQILEKLPNKVEIDGKYVYVYDNGKVRNFTLIADGKILHDETGAYTDSIDTVNKDI 211
QY 225 TTNIVDGFSINNRAYDSSEASPELIDGYLTADSWTRPASIIKQGVTWQASTAEDFPPLM 284
Db 212 VTTRSNLYKKNQVYDRSAQSFHFVDHYLTAEBSWYRPKYILKDKGTWTQSTEKDFPLM 271
QY 285 AWWPNVDQVNYLNYMVKVFNLDKAYSSTDKOETLKVAAKDIQIKI EOKIOAEKSTOWLR 344
Db 272 TWVPOETQORQVNTVMNAQLGINKTYDDTSNQLNLNIAAATIOAKIEAKITLKNTDWLR 331
QY 345 ETISAFVKTQPMWNETENYSGGGEHLQGGALLYVNDNR--TPWANSYRRLNRTATNQ 403
Db 332 QTI SAFVKTQSAWNSDSEK---PFDDHLQNGAVLYDNEBKLTTPVANSNYRILNRTPTNQ 387
QY 404 TGTIDKSLIDEGSDP-----NHMGGFDFLLANDVLSNPVQAEQLNQIHYLMNWSIVM 458
Db 388 TG-----KQDPRTADNTIGGYEFLLANDVDNSNFPVQAEQLNLWHLFNLNFGNIYA 438
QY 459 GPKDANFDGIRVDADVNDADMLQLYTNYFREYVGVNKSSEANALAHISVLEAWSLNDHY 518
Db 439 NDPDANFOSIRVDADVNDADLLQTAGDYLKAAGKIHKNDKAANDHLSILEAWSNDTPY 498
QY 519 NDKTDGAALAMENKORLALLFSLAKPIKERTPAVSPLYNNFT--NTTORDEKTDWINKGS 577
Db 499 LHDDGDNMINMDKRLSLFSLAKPLNQRS--GMNPLITNSLVNRTDDNAET-----549
QY 578 KAYNEDGTVQSTIGIKYNEKYGDASGNVYFIRAHNNVQDIIAEIIEKEINPKSDGFTIT 637
Db 550 -----AAVPYSYFIRAHDSVQDLADIIFKAEINPNVVGYSFT 587
QY 638 DAEMKQAPEIYNKMLSSDKKTYTLANNIPAAVAMLQNMETITRVYVYDGLYDGDGHMETK 697
Db 588 MEEIKKAFEIYNKDLLEATEKKYTHNTALSYALLLNKSSVPRVYVYVYVYVYVYVYVYVYV 647
QY 698 SPYDITVNLMSRLKIYVSGGQAQSYWLPDQKMDNSDVELYRTNEVYTSVRYKDIWAND 757
Db 648 TINYEAETLLKARIKYVSGGQAMRQ-----QVGN-----EIIITSVRYGKALK 693
```


Db 770 TTDNGIKAYHSDQ--AAGLVRYTNDRGELIFTAADIKGANPQVSGYLGVWVPVGAAD 827
Qy 869 QDIRVAPSTBAKEGEELTLKATEAYDSOLIEGFSNFQTI PDGSDPSVYNNRIKAENVDL 928
Db 828 QDVRVAASTAPSTDGK-SVHQNAALDSRVWFEGFSNFQAF--ATKKEEYTNVIAKNVDK 884
Qy 929 FKSQGVTSFEMAPQFVSADDTGLSDSVIONGYAPADRYDLAMSKNNKYSGKEDLRDALKA 988
Db 885 FAEWGVTFEMAPQVSVSDGSLSDSVIONGYAFTRDYDLGSKPNKYGTADDLVKAICA 944
Qy 989 LHKAGIOAIADWPDQIYOLPGKEVVTATRTDAGRKIAADAIIDHSLYVANSKSSGKDIQ 1048
Db 945 LHSKGIKVMADWPDQIYALPEKEVVTATRTDAGRKIAADAIIDHSLYVANSKSSGKDIQ 1048
Qy 1049 AKTGGFLELAKKAKYPMFKVMNISTGKPIDDSVKLWKMAEYFNGTNNVLERGVGVLSD 1108
Db 1005 AKYGGAFLELQAKYFELFARKQISTGVNPDPSVKIKQWAKYFNGTNNVLERGAGYVLKD 1064
Qy 1109 EATGKYFTVTKGNFPLQLT----GKEKVTGSSDGKGYITGTSQTQAKSAFVTFNG 1164
Db 1065 QATNTYFSLVSDNTFLPKSLVNPNGHGTSSVTGLVFDGKGYVYVYSTSGNQAQNAFISLGN 1124
Qy 1165 NTYYEDARGHMVTNSEYSPNGKDVYRFLPNGIMLSNAPYIDANGTYL--YNSKGQMYKG 1222
Db 1125 NWTYFDNNGYMTGAQ-SINGANY-PLSNGIQLRNAIY--DNGNKVLSYTGNDGRRYEN 1180
Qy 1223 GYTKFDVSETDKDGKSKVVKPFYFTNEGVMAGVTVTIDGFTQYFEGEDGFOAKDLV-TF 1281
Db 1181 GYVLF-----GQQ-----WYFON-GIMAVGLTRVHGAOYFQASGFQAKQGFITTA 1226
Qy 1282 KGTYTFDAHTNGIKIDTW-RNINGKWTYFPDANGVAATGAQVNGOKLYFNEPDGSOVKGG 1340
Db 1227 DGKLYFDRDSGNQISNRFVRNSKGEWFLFDHNGVAVTGVTENGQRLYFKPKNVQAKGE 1286
Qy 1341 VKNADGTYKYEKGEGELVTNEFFTTDGNVWYTAGANGTKVTGQAOVING 1390
Db 1287 FIRDANGYLYRYPDPSNGEYRNRFRVNSKGEWFLFDHNGIAVGTGARVWG 1336

RESULT 9

US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans

US-09-740-274-4

Query Match

38.7%; Score 3070; DB 2; Length 1375;

Best Local Similarity 46.2%; Pred. No. 1.5e-193;
Matches 661; Conservative 223; Mismatches 412; Indels 134; Gaps 33;
Qy 1 MENKIHYKHLKHKKOWWTTAVASVALATVLGGISVTTSSVSADETODK--TVTQSNSTGT 58
Db 1 MEKKVAFKURKKRWVTVSIAS-AVVTL---TSLGSLVKADSTDDROQAQTESQ---- 52
Qy 59 ASLVTSPETAKEADKRTNTKEADVLPKAKETNAVETATTTNTQATAEAATATTADVAVA 118
Db 53 ASLVTTSEAAKETLTATDITSTATSATSQPTATVDNVSTTN-QST---NTTANTANFVK 108
Qy 119 AVNKEAVVTDA-PAVTTEKABEQ-----PATVKAENVNTEVKAPE-- 159
Db 109 PTTTSEQAKTDSDKIITTSKAVNRLTATGKVPANNNTAHPKTVTDKIVIPKPKIGK 168
Qy 160 --AALKSDSEVEAALSUKNKNIDGKYVYVNEGSHKENFAITVNGQLLYFGKQALTS 217
Db 169 QPSLSQDDIAALGNVKNIRKVGKYYKYKEDGTQKNVALNNGTKTFFDFETGALSNT 228
Qy 218 TYSFTPTGTTN--IVDGFSSINNRAYDSSEASFELIDGTLTADSWYRPAASIKDGV 275
Db 229 LPSKKGNIINNDNTNSFAQYVQVYSTDVANFEHVDHYLTABSWYRKYILKDKGT 288
Qy 276 AEDFRPLMAWPNVDTVNYLNYSKVNFLDAKYSTDKQETLKVAAKDIQIKIBOKIQ 335
Db 289 EKDFRPLMTWPDQETQRYVYNNYAQLGIHOTYNTATSPQLQNLAAQTIQTKIEKIT 348
Qy 336 AESTOMLARETISAFVKTQPNWNETENYSGGEGEDHLOGGALLYNDYR-TPWANS 394
Db 349 AEKNTWMLRQTIISAFVKTQSANNSDSEK----PFDDHLOKGLLYNNNSKLT 404
Qy 395 RLNRATNTGTITDKSILDEQSDPNH-----MGFPDFLLANDVDLSNPVVAQALMOIHY 449
Db 405 ILNRFTNQTG-----KKDPRTABRTICGYEFLANDVDNSNPVVAQALMLWLF 455
Qy 450 LAMWGSIVMGDKDANPDGIRVDADVNDADMLQLYTNYFREYGVNKNSEANALAHISVLE 509
Db 456 LAMFGNIYANDPDANFDSIRVDADVNDADLLQIAGDYLKAAKGIHKNDKCAANDHLSILE 515
Qy 510 ANSLADNHNNDKTDGAALAMENKORLALLFSLAKPIKERTPAVSPLYNNTF-NTTORDEK 568
Db 516 AWSNYDTPYLHDDGDNMINMDNRLRLSLYSLAKPLNQRS-GMNPLITNSLVNRTD 574
Qy 569 TDWINKDGSKAYNEDGTQVKTIGKNEYKGASGVYFIRAHNNVDQIIAIIKKEIN 628
Db 575 T-----AAVPSYSFIRAHDSVQDLIRNIIITEIN 604
Qy 629 PKSDGFTITDAEMKQAFIYNKMDLSSDKKYLNNIPAAAYAVMLQNMETITRYVYGDLYT 688
Db 605 PNWGVSYFTTEBIKKAFIYNKDLLATEKKYTHYNTALSYALLTNKSSVPRVYVYGD 664
Qy 689 DDGHYMETKSPYDITVNLKMSRIKYVSGQQRSTWLPDGDGMNDSDELVELYKTN 748
Db 665 DDQYMAHKTINYEAITELLKARIKYVSGQAMRNQ-----QVGNSS-----BIITS 710
Qy 749 VRYGKDIIMTANDTEGSKYSRTSQVTLVANPNKLNLDQSAKLVNKGKTHANOKYRALIV 808
Db 711 VRYGKALMATDT-GDRTTTSQVAVIEGNNSLRLKASDRVVVNVNNGAAHKNQAYPL 769
Qy 809 GTADGINKFTSDADAIAAGYVKETDSNGVLTFCANDIKGYETDFDMSGFVAVVPV 868
Db 770 TTDNGIKAYHSDQ--AAGLVRYTNDRGELIFTAADIKGANPQVSGYLGVWVPVGAAD 827
Qy 869 QDIRVAPSTBAKEGEELTLKATEAYDSOLIEGFSNFQTI PDGSDPSVYNNRIKAENVDL 928
Db 828 QDVRVAASTAPSTDGK-SVHQNAALDSRVWFEGFSNFQAF--ATKKEEYTNVIAKNVDK 884
Qy 929 FKSQGVTSFEMAPQFVSADDTGLSDSVIONGYAPADRYDLAMSKNNKYSGKEDLRDALKA 988
Db 885 FAEWGVTFEMAPQVSVSDGSLSDSVIONGYAFTRDYDLGSKPNKYGTADDLVKAICA 944
Qy 989 LHKAGIOAIADWPDQIYOLPGKEVVTATRTDAGRKIAADAIIDHSLYVANSKSSGKDIQ 1048

Db 945 LHSKGIKWADVPDQMYALPEKEVVTATRVVDKYCTPVAGSQIKNTLYVVDGKSSGKDOQ 1004
Qy 1049 KYGGSEFLAELKAKYPMFKVNMISTGKPIDDSVKLKQWKAERYFNGTNNVLERGVGVLSLD 1108
Db 1005 KYGGAFLELQAKYPELFARKQISTGVPMDSVKIKQWSAKYFNGTNIILGRGAGYVLKD 1064
Qy 1109 EATGKYFTVTKEGNRTPLOLT---GKEKVIITGSSDGGKITYFGTSGTQAKSAFVTFNG 1164
Db 1065 QATNTYFSLVSONTEFLPKSLVNPNGHTSSSVTGLVDPDGKGYVYVTSNGQAKNAFISGN 1124
Qy 1165 NTYFDFARGHMTNSESYPNGKDVYRFLPENGIMLSNAFYIDANGNYL---YNSKGMQYKG 1222
Db 1125 NYFYDNNGYMTGAQ-SINGANY-FLSNGIQLRNIY--DNGKVLSSYNGDGERYEN 1180
Qy 1223 GYTKFDVSETDKGKESKVVKRYFTNEGVMKAGVTVIDGTYQYFGEDEGQAKDLV-TF 1281
Db 1181 GYLYF-----GQQ-----WRYFON-GIMAVGLTRVHGAQVYFDASGFAKGQFITTA 1226
Qy 1282 KGTYTFDAHTNGIKDTH-RNINGKMYFDFANGVNAATGAOVINGOKLYFNEDEGSOVKG 1340
Db 1227 DGKLYFDRDSGNQISNRFRNSKGEWFLPDHNGVAVTGTVTFNGQRLYFKPNGVOAKGE 1286
Qy 1341 VKNADGTSKYKGBGEGELVTNEFTTDCGNVYVYAGANGKTVTGAOVING 1390
Db 1287 FIRDANGILRYDPSNGNEVRNFRVNSKGEWFLPDHNGIATGARVNG 1336

RESULT 10
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match 29.4%; Score 2327; DB 2; Length 2057;
Best Local Similarity 38.2%; Pred. No. 2.9e-144;
Matches 577; Conservative 209; Mismatches 509; Indels 217; Gaps 49;
Qy 144 ATVKAERVNTEVKAPPAALKDSEVEAALS-----LKNIKNIDGKYYVN 187
Db 300 AVANTEIINGKLQGRDA--NGHQVNAFSKDVAGNTFFPDANGVMLTGLQISGKTYILD 357
Qy 188 EDGSHKENFAITVNGQLLYFGKD-GALTSSTSYSPFGPTTNIVDGFSINNRAYDSEASF 246
Db 358 EQGHLRKNYAGTFNNGFMFYFADTGAKTAIEYQPDQGLVSSQSNENTPHNAKSYDKSSF 417
Qy 247 ELIDGVLTDASVYRPASIIKDGVTWQASTAEDRPLLMAMWPNVDTQVNYLYNM-SKVEN 305
Db 418 ENVDGYLTADTVRYPDILKNGDTWTASTEDMRPLMTWPDQKQANLYNFMSSKGLG 477
Qy 306 LQAKSYSTDKQETLKVAAKDIOIKIEQKIOAKESTOWLRETTISAFVKTOPQWNKETENYS 365
Db 478 ITTYTAASQKTLNDAAFVQTAIEQOISLKSTEWLDAIDSFVKTOANWKNQTEDEA 537
Qy 366 KGGEDHLGGALLVYND-S-RTPWANS-DYRRLNRATATQGTIDKSIDLEQSDPNHMG 423
Db 538 FDGLQ-WLQGGFLAYQDDSHRTPNTDSSGNNRKLGRQPIN-----IDGSKDTTDGKG 587
Qy 424 FDFELLANDVLSNPVQAEQLNIHLYLMNWGSIWMDKANDPDGIRVDADVNDADMLQL 483

Db 588 SEFLLANDIDNSNPVQAEQLNHLNMFSSITGNNDNANPDGIRVDADVNDADMLKI 647
Qy 484 YTYPREYGVNKSANALAHISLEAWSLNDNHNNDKTDGAALAMENKQRLALLFLSLAK 543
Db 648 AGDYFKALYGTDKSDANANKHLSILELDWNGKDPQVYNOQNAQLTMDY-----695
Qy 544 PIKERTPAVSPLYNNTFTTQORDEKTDW-----INKDGSKAYNEDGTGVKQSTIGKN 595
Db 696 -----TVTQFGNSLTHGANNRNNMYFLDTGYLYLNGDLNKKIIVDKNRPNSSGLTVNRI 748
Qy 596 EKYGDAS--GNVYFTRAHDNNVDII--AEIIKKEINPKSDGFTTDAEMKQAFETYNKD 651
Db 749 ANSGDTKVIPNYSFVRANDYDAODPIRKAMIDHGIKNMQDTFTFD--QLAOCMEFYKD 806
Qy 652 M--LSSDKKTYLNNIPAAAYAVMLQMETITRYYGDLTDDGHYMETIKSPYYDTIVNLMK 709
Db 807 QENPSGFKYNDYLPAYAMLLTNKDTVPRVYGYDMYLEGGQYMEKGTIYNPVISALLK 866
Qy 710 SRIKTVSGGQARQSRVLTPTDGKMONSDVELYRTNEVYTSVRYGKDIIMTANDT---EGSKY 766
Db 867 ARIKTVSGGTQMAT---DSSGK-DLKDGE---TDLTTSVRFGKGMTSDQTTTQONSOD 918
Qy 767 SRTSGOVLTVANNPKLNDQSAKLVNEMGKIHANOKYRALIVGTADGIKNFTSDADIAA 826
Db 919 YKQIGIVIGVGNPDLKLNNDKTIITLHMGAHKNQYRALVLSNDSGIDVDYDDKAPT 978
Qy 827 GYVKETDSNGVLTFGAN-----DIKGYETFDMSGFVAVVVPVGVASDNDQDIR- 872
Db 979 ---RTNDNGDLI FHKTNTFVKQDGTIINYEKMGSLNALISGVLGVVVPVGVASDSQDART 1034
Qy 873 VAPSTEAKKEBELTLKATEAYDSQLIYEGFSNFQIIPDGSDFSVVTRNRIKIAENVDLFKSW 932
Db 1035 VATESSSSDGVS- PHSNAALDSNVIYEGFSNFQAMP--TSPEQSTNVVVIATKANLFKSL 1091
Qy 933 GYVTSFEMAPOFVSADDG-----TFLDSVIQNGYAFADRYDLAMSKNN-----KYGSKED 981
Db 1092 GITSFELAPQNRSSGDTNYGMSFLDSFLNNGYATDRYDLGFKADGNPNPKYGTQDQ 1151
Qy 982 LRDALKALHKGAGIAIADWVPDQIYQLPFGKEVVTATRTDGGARKIADAIIDHSLYVANSK 1041
Db 1152 LRNAIEALHKGQMAIADWVPDQIYALPFGKEVVTATRVDERGNQLKDTDFVNLLYVANTK 1211
Qy 1042 SSGKDYQAKYGGEFELAEIKAKYPEMFKVNMISTGKPIDDSVKLKQWKAERYFNGTNNVLERG 1101
Db 1212 SSGVDYQAKYGGEFELAEIKAKYPEMFKVNMISTGKPIDDSVKLKQWKAERYFNGTNNVLERG 1271
Qy 1102 VGVLSDATGKYFTVTKEGN-FIPLQLTGKEKVIITGFSDDGKIGITFGTSGTQAKSAFV 1160
Db 1272 AYYVLKDWATNQYFNIKTNEVFLPLQONKD-AQTGFISDASGVKYISISGQAKDTFI 1330
Qy 1161 -TFNGNTYFDFARGHMTN-----EYSPNGKD-VYRFLPENGIMLSNAPYIDANGN 1209
Db 1331 EDGNGMYVYFDRKGYMVRSSQGENPIRTVETSVNTRNGNYFPMVNGVELRKGFGTDNSGN 1390
Qy 1210 TYLYNSKQMYKGGYTKFDVSETDKGKESKVVKRYFTNEGVMKAGVTVIDGFT-QYFG 1268
Db 1391 VYFPDQGGKVRDKYINDAN-----NPFHLNVDMGTMRGLFKFSDTLQYFA 1438
Qy 1269 EDGFOAKDLV-TFKGTYYPDAHTGN--GIKDYTRNNGKMYF-----DANGVAATGA 1320
Db 1439 SNGVQIKDSYAKDSKNGKYFDSATGNNDTGKAQTH---DNGGYITIDSDANNITGVNT 1495
Qy 1321 QVINGQKLYFNEBDG---SQVKGGVVKNADG-----TYSKYKEGFGELVTFNEPFTDQ 1369
Db 1496 DYTAVITSSLEDGLFANAPYGVVTKDQNGNDLKWQYINHTKQYEGQQVQVTRQYTDKSG 1555
Qy 1370 NWV-----YYAG-----ANGKTVTCAQVIN-----GOHLYFNADGSOVKG- 1404
Db 1556 VSWNLITFAGDLOGQRLWDSRALTPFKTMNQISFISYANRNDGLFLNAP-YGVKGY 1614
Qy 1405 -----GV--VKNADGTVSKYNAS---TGERLUTNEFTTGNNNWYIG- 1441

Db 1615 QLAGSMNOYKGOQVTTIAGVANVSCKOWSLISFNGTQVWIDSQALNTNFTHDMNQKVFVNT 1674
Qy 1442 -ANGKSVTGEVKIGDDTYFFAKDKQVKQGVTSAGNGRISYIYGYDSGKRAVSWIE-IQP 1499
Db 1675 TSNUDLGLFLNAPYRQPCYKLAGLAKNNTQVTVS-----QQYFDDQG-----TWVSQVVLG 1726
Qy 1500 GVVYVFDKNGLA 1511
Db 1727 GQTVWVDNHALA 1738
RESULT 11
US-09-604-957-3
; Sequence 3, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUTI, HAKIM
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-3
Query Match 26.4%; Score 2093; DB 2; Length 1278;
Best Local Similarity 37.2%; Pred. No. 3.8e-129;
Matches 511; Conservative 197; Mismatches 414; Indels 250; Gaps 42;
Qy 18 TIASVASVALATVLGGLSVT---TSSVSADETQKVTQNSGTTASLVTSPKATK----- 69
Db 54 TIKLTDQYQALNGQLVLLRFSKAADGNPSGNTVTDQFSKRYATGTGDFVYKNGNQ 113
Qy 70 -----EADKRWTKBAD---VLTPAKETNAVETATTTNTQATPAEATATTADV--AVA 118
Db 114 VEPFGWHATQNSDKDSQIIVLVNGKE---VKRQLVNDTK---EGAAGFNRRNDVYKVP 167
Qy 119 AVPNK-----EAVVTDPAVTEKABEQATVKAERV-----NTEVKAPER----- 160
Db 168 AIENSSMSGFQGIITL-----PVTYKNNVQLVHRFSNDVKTGEGNYVDVFS 214
Qy 161 ---ALKDSVEEAALSILK--NIKNIIDGKYVYVN-EDGSHKENFAITVNGQLLYFGKD-GAL 213
Db 215 ELPVVDKSKNGPLKQGLQTINGQQYIYDPTQPRKFLQSGNNWIYFDSGTGVG 274
Qy 214 TSSSTVSFTPGTNNIYVDGFSINNRAYDSSEASFELIDGYLTADSWYRPAIIKDGVTWQA 273
Db 275 TWALELQPAKGTVSSNEQYRNGNAAYSYDDKSIENVNGVLTADTWYRPAIILKDGTTGTTD 334
Qy 274 STAEPRPLMAWPNVDVQVYLVNTMSK-----VFNLDAKYSSTDQKQETLKVA 323
Db 335 SKETDMRPILMYWPNPNTLTQAYLYNTMKQHGNLLPSALPFFNADA-----DPAE-LNHYS 388
Qy 324 KDQIKIEOKIQAESTQWLRITISAFVKVTOQWNNKTESYKGGEDHLOGGALLYVND 383
Db 389 EIVQONIEKRISSETGNDMLRLMHDFTVNNPMNKSSENVPFSGIQ--FQGGFLKYENS 446
Qy 384 SRTPWANSYRRLNRTATNQTGTIDKSIILDEQSDPNHMGGFDFLLANDVLDNSPNVVAEQ 443
Db 447 DLTYPANSYRLLGRMPIN-----IKDQT-----YKQGFLLANDIDNSPNVVAEQ 493
Qy 444 LNIHLYLMNWSGIVGDKDANFDGIRVADVNDVMDLQLYTNYFREYYGVNKSANAL 503
Db 494 LNWLYLLNFGTITANNQANFDSVRVDPADPNIDALMNIADQYFNAAYGMD-SDAVSNK 552

Qy 504 HISVLEAWSLNDNHYNDKTDGAALAMENKORLALLFSLAKPIKERTPAVSPLYNNTNTT 563
Db 553 HINILEDWHADEYFNKIGNPOLTWDDTIKNSLHGLS----- 591
Qy 564 QRDEKTDWINKOGSKAYNEBGTQVQSTIGKYNKYGD-ASGNVYFIRAHNNVQDIIAEI 622
Db 592 -----DATNRWGLDAI VHQSILADRENNSTENVVIPNYSFVRAHDNNSQDQIQNA 640
Qy 623 IKKEINPKSDGFTITDAEMKQAFEIYNKDWLSDDKYYTLNNIPAAVAMLMONMETITRVY 682
Db 641 I-RDVTGK--DYHTFTFEDBQGDIDAYIQDNSTVKKYNLYNIPASVAILTNNKDTIPRVY 698
Qy 683 YGDLTYDDGHYMETKSPYYDTIYNLMKSRIKYVSGGAORSYWLPTDGRKMDNSDVELYRT 742
Db 699 YGDLTYDDGGQYMEHQRYVDTLTNLLKSRVYVAGGQSMQT-----MSVGGN 745
Qy 743 NEVTSVRYGKOIMTANDTEGSKYRTSGQVTLVANNPKLNLDQSKLANVEMGKIHANQK 802
Db 746 NNILTSVRYGKGAMTATDT-TGDETRTQIGVGVVSNTPNLKLGVDNKVVLHMGAAHKQKQ 804
Qy 803 YRALIVGTADGIGNFTSDADATAAGVYKETSNGVLTFGAND-----IKGYET 850
Db 805 YRAAVLTITDGVINTSDQGPVA---MTDENGDLYLSSHNLVNVNGKEEADTAVQGYAN 860
Qy 851 FDMGSGFVAVVVPVVGASDNODIRVAPSTEAKKEGELTLKATEAYDSOLLYEGRESNFQITPD 910
Db 861 PDSVGYLAVVVPVVGASDNODARTAPSTE-KNSGNSAYRTNAAPFNSVIFAEFNFVYTP- 918
Qy 911 GSDPSVYTNRKIAENVDLFKSMGVTSFEMAPQVFSADDTGFLDSVQNGYAFADRYDLAM 970
Db 919 -TKESERANVRIAQNADPFASLGTFSFEMAPQVYNSKDRTFDSTIDNGYAFDTRVDLGM 977
Qy 971 SKNKYKSGKEDLRDALKALHAKAGIQAIADWDPDQIYQLPGKEVTVTRTDGAGRKIADAI 1030
Db 978 SEPNIKYGTDEDLRNAIQALHAKAGLQVMAWDPDQIYNLPFGKEVATVTRVDRGNVWKDAI 1037
Qy 1031 IDHSLYLVANSKSGDKYQAKYGGEFLEALKAKYPEMFVKVNMISTGKPIDDSVKLKQWKAE 1090
Db 1038 INNLYNVNVTICGG-EYQKYGAFDLQKLYPEIFTKKQVSTGVVAIDPSQKITEWSAK 1096
Qy 1091 YFNGTNVLRGCVYVLSDEATGKYF---TVTKEGNFIPQLTGKKEKVIITGFSDDGKITY 1147
Db 1097 YFNGTNILHRGSGVYLKADG-QQYVNLGTTTQ--FLPLQLTGEK- 1139
Qy 1148 FGTSGTQAKSAFTFN-GNTYVFDARGHMVNTSEYSPNGKDVYRFLPNGIMLSNAYIDA 1206
Db 1140 -----QGNFVFGKDGNYFYFDLAGNMV-----KNTFIEDS 1171
Qy 1207 NGNTYLYNSKGQYKGGYTKPDVSETDKDGKSKVVKPRYFTNEGVMAGVTVIDGF--- 1263
Db 1172 VGNWTF- 1199
Qy 1264 -TQYFEGEDGFQAKDLVTFKGYTYFADATNGNGIKDTRNNGKWKYVFDANG 1314
Db 1200 GTYFELKNGVSRGGLVQTDNGTYFDNY-GHGVENQITINAGAMITYTLDENG 1250

RESULT 12

US-09-995-749A-2
; Sequence 2, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUTI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28

Qy	851	FDMSGFVAWVPGASDNQDIRVAPSTEAKKEGELTKATEAYDSQLIYEGFSNFQTPD	910
Db	1364	PDVSGYLAWVPGASDNQDARTAPSTE-KNSGNSAYRTNAAPDSNVIFEAFSNFVYTP-	1421
Qy	911	GSDPSVYVNRKKAENVDLFKSMGVTSEFEMAPQFVSADGTFDLSVTLQNGYAFADRYDLAM	970
Db	1422	-TKESERANVRIAQNADPFASLGFTSFEMAPQYNSSKDRFLDSTIDNGYAFTRDYDLGM	1480
Qy	971	SKNNKYGSKEDRLDALUKHAGIAIADVPDQIYQLPGKEVYVTRATRDGAGRKIDAI	1030
Db	1481	SEPNKYGTDELRNIAQALHKAGLQWADWPDPQIYNLPQKEVATVTRVDDRGNVWKDAI	1540
Qy	1031	IDHSLVANSKSGKDYQAKYGGEFLLAEKAKYPEMFKVMNISTGKPIDDSVKLKQWAE	1090
Db	1541	INNLLVAVNTIGGG-EYQKYGGAFLDKLQKLYPEIFTKKQVSTGVAIDSPQKITEWSAK	1599
Qy	1091	YFNGTNVLRGVGYVLSDAETGKYF--TVTKEGNFIPQLQITGKEVITGFSSDGGKIT	1147
Db	1600	YFNGTNILHRGSGYVLKADG-QQYNNLGGTTTKQ--FLPIQLTGEK--	1642
Qy	1148	FGTSGTQAQSAFVTFN-GNTYFDFARGHVTNSEYSPNGKDVYRFLPNGIMLSNAFYDA	1206
Db	1643	-----QGNEGFVKNGDGNIFYDYDLAGNMY-----KNTFTIEDS	1674
Qy	1207	NGNTLYNSKGOMYGGYTKFDVSESTDQDGESKVKVFRYFTNEGVMAGVTVIDGF--	1263
Db	1675	VGNWYFF-----DQDG---KMVENKHFDV-----VDSYGEK	1702
Qy	1264	-TOYFGEDGFQAKDLVTPFKTYTFDAHTGNGIKDTRWNINGSKWYTFDANG	1314
Db	1703	GTVFYFLKNGSVSRGGLVQTDNGTYYPDNY-GKMVRNQTINAGAMIYTLDENG	1753
RESULT 13			
US-09-604-957-4			
; Sequence 4, Application US/09604957			
; Patent No. 6486314			
; GENERAL INFORMATION:			
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA			
; APPLICANT: DLJKHUIZEN, LUBBERT			
; APPLICANT: RAHAOUI, HAKIM			
; APPLICANT: LEER, ROBERT-JAN			
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN			
; FILE REFERENCE: BO 43388			
; CURRENT APPLICATION NUMBER: US/09/604,957			
; CURRENT FILING DATE: 2000-06-28			
; PRIOR APPLICATION NUMBER: 00201871.1			
; PRIOR FILING DATE: 2000-05-25			
; NUMBER OF SEQ ID NOS: 17			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 4			
; LENGTH: 545			
; TYPE: PRT			
; ORGANISM: Streptococcus mutans			
US-09-604-957-4			
Query Match 20.0%; Score 1584.5; DB 2; Length 545;			
Best Local Similarity 53.9%; Pred. No. 3.6e-96;			
Matches 316; Conservative 100; Mismatches 121; Indels 49; Gaps 10;			
Qy	427	LLANDVDLNPVVOAEOALNQIHYLMKNWGSIVMGCKDANPDGIRVDADVNDADMLQLYTN	486
Db	1	LLANDIDNSNPVVOAEOALNWLHYLMYGSIVANDPEANFDGVRDADVNNADLLQITASD	60
Qy	487	YFREYYGVNKSSEANALAHISVLEAWSLNDNHYNDKTDGAALAMENKORLALLFSIAKPIK	546
Db	61	YLKAHYGVDKSEKQANHLHSLEAWSNDPQYNKDTKGALPIDNKLRLSILLYALTRPLE	120
Qy	547	-----ERTPAVSPLYNNFTNTQORDEKTDWLNKDGSKAYNEBDGTVKQSTIGKYNKYG	599
Db	121	KDASNKNEIRSGLEPVITSLN-----NRSAGENKSR--	153

QY 600 DASGNVFIIRAHNNVQDIIEAIIKKEINPKSDGFTITDAEMKQAFIYNKQMLSSDKKY 659
Db 154 --MANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFLDELKQAFKIYNEDMRQAKKY 211
QY 660 TLANNIPAAVAVMLQNNMETITRVYGYDLYTDDGHYMETKSPYDTIVNLMKSRIKYSGGQ 719
Db 212 TQSNIPAYALMLSNKDSITRLYGYDMSYSDGQYMATKSPYYDAITLLKARIKYAAGGQ 271
QY 720 AQRSYMLPTDGMNDSDVELYRTNEVYTSVRYGKDIINTANDTEGSKYSRTSGQVTLVANN 779
Db 272 DMKITTV--EGDKSHMD--WDVTGLTSSVRYGTGANEATD-QGSEATKTQGMVITSNN 325
QY 780 PKLNLQDSAKLVNEMGKIHANOKYRALIVGTADGINKFTSDADAIAAGYVYKETSNGVLT 839
Db 326 PSLKLNQNDKVIYVMGAHKNQYRPLLTTKDGLTSYTS--DAAAKSLYRKTNDRGELV 383
QY 840 FGANDIKGYETFDMSGFVAWVPVPGASDNQDIRVAPSTEAKKEGELTLKATEAYDSOLLY 899
Db 384 FDSADLQGLYNLPQVSGYLAVWVPVPGASDNQDVRVAASNKANATGQV-YESSALDSOLLY 442
QY 900 EGSFNFQT-IPDGSFSPVYTNRKIAENVDLFKSWGVTSPFEMAPQFVSADDTFLDSVION 958
Db 443 EGSFNFQDFVTIKDSD--YTNKKIAQNVQLFKSWGVTSPFEMAPQYVSSSDGSFLDSIIQ 499
QY 959 GYAFADRYDLAMSKNNKYSGKEDLRDALKALHKGAGIQAIDWVPDQ 1004
Db 500 GYAFEDRYDLAMSKNNKYSGQODMINAVKALHKSGLQVIADWVPDQ 545

RESULT 14

US-09-995-749A-10

; Sequence 10, Application US/09995749A

; Patent No. 6867026

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAOUI, HAKIM

; APPLICANT: LEER, ROBERT-JAN

; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES

; FILE REFERENCE: BO43388-CIP

; CURRENT APPLICATION NUMBER: US/09/995,749A

; CURRENT FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: 09/604,957

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: EPO 00201871.1

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 10

; LENGTH: 545

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-09-995-749A-10

Query Match 19.7%; Score 1560.5; DB 2; Length 545;
Best Local Similarity 53.8%; Pred. No. 1.4e-94;
Matches 316; Conservative 99; Mismatches 121; Indels 51; Gaps 12;

QY 427 LLANDVDLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDVADVNDMLQLYTN 486
Db 1 LLANDIDNSNPVQAEQLNWLHYLMNYGSIVANDPEANFDGVRVDVADVNDMLQIAD 60
QY 487 YPREYGVNKSANLAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLFSLAKPIK 546
Db 61 YLKAHGVDSKSNAINHLSILEAWSNDNPQYKDTKGALQFLDNKRLSLLYALTRPUE 120
QY 547 -----ERTPAVSPLYNNFTNTORDEKTDWINKDGSKAYNEDGTGVTQSTIGKNEYKYG 599
Db 121 KDSASGNKRSGLPEVITNSLN-----NRSAGKNSER-- 153
QY 600 DASGNVFIIRAHNNVQDIIEAIIKKEINPKSDGFTITDAEMKQAFIYNKQMLSSDKKY 659
Db 154 --MANYIFIRAHDSVQTVIAKIIKAQINPKTDGLTFLDELKQAFKIYNEDMRQAKKY 211

QY 660 TLANNIPAAVAVMLQNNMETITRVYGYDLYTDDGHYMETKSPYDTIVNLMKSRIKYSGGQ 719
Db 212 TQSNIPAYALMLSNKDSITRLYGYDMSYSDGQYMATKSPYYDAITLLKARIKYAAGGQ 271
QY 720 AQRSYMLPTDGMNDSDVELYRTNEVYTSVRYGKDIINTANDTEGSKYSRTSGQVTLVANN 779
Db 272 DMKITTV--EGDKSHMD--WDVTGLTSSVRYGTGANEATD-QGSEATKTQGMVITSNN 325
QY 780 PKLNLQDSAKLVNEMGKIHANOKYRALIVGTADGINKFTSDADAIAAGYVYKETSNGVLT 839
Db 326 PSLKLNQNDKVIYVMGAHKNQYRPLLTTKDGLTSYTS--DAAAKSLYRKTNDRGELV 383
QY 840 FGANDIKG-YETFDMSGFVAWVPVPGASDNQDIRVAPSTEAKKEGELTLKATEAYDSOLLY 898
Db 384 FDSADLQGLYNLPQVSGYLAVWVPVPGASDNQDVRVAASNKANATGQV-YESSALDSOLLY 441
QY 899 EGSFNFQT-IPDGSFSPVYTNRKIAENVDLFKSWGVTSPFEMAPQFVSADDTFLDSVION 957
Db 442 EGSFNFQDFVTIKDSD--YTNKKIAQNVQLFKSWGVTSPFEMAPQYVSSSDGSFLDSIIQ 498
QY 958 GYAFADRYDLAMSKNNKYSGKEDLRDALKALHKGAGIQAIDWVPDQ 1004
Db 499 GYAFEDRYDLAMSKNNKYSGQODMINAVKALHKSGLQVIADWVPDQ 545

RESULT 15

US-09-604-957-5

; Sequence 5, Application US/09604957

; Patent No. 6486314

; GENERAL INFORMATION:

; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

; APPLICANT: DIJKHUIZEN, LUBBERT

; APPLICANT: RAHAOUI, HAKIM

; APPLICANT: LEER, ROBERT-JAN

; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN

; FILE REFERENCE: BO 43388

; CURRENT APPLICATION NUMBER: US/09/604,957

; CURRENT FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: 00201871.1

; PRIOR FILING DATE: 2000-05-25

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 5

; LENGTH: 523

; TYPE: PRT

; ORGANISM: Leuconostoc mesenteroides

US-09-604-957-5

Query Match 15.9%; Score 1262.5; DB 2; Length 523;
Best Local Similarity 46.5%; Pred. No. 5.9e-75;
Matches 269; Conservative 84; Mismatches 170; Indels 55; Gaps 8;

QY 427 LLANDVDLSNPVQAEQLNQIHYLMNWSIVMGDKDANFDGIRVDVADVNDMLQLYTN 486
Db 1 LLANDVDNSNPVQAEQLNWLHYLMNFGTITANDANFDGIRVDVADVNDMLQIAD 60
QY 487 YPREYGVNKSANLAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLFSLAKPIK 546
Db 61 YKFLAYGVQNDATANHLSILEAWSNDPLVYTDQSGNQLTMDDDYVHTQLIWSLTSSD 120
QY 547 ERTPAVSPLYNNFTNTORDEKTDWINKDGSKAYNEDGTGVTQSTIGKNEYKYGDSGNV 606
Db 121 IR-----GTMRQFVDYVWVDRSNDSTENE-----ALPNYS 150
QY 607 FTRAHDNNVQDIIEAIIKKEINPKSDGFTITDAEMKQAFIYNKQMLSSDKKYTLNNIPA 666
Db 151 FVRAHDSVQTVIAQIVSDLYPDVENS LAPTEQLAAAFKVYNEDEKLAKKYQYNMAS 210
QY 667 AYAVMLQNNMETITRVYGYDLYTDDGHYMETKSPYDTIVNLMKSRIKYSGQVTLVANN 726
Db 211 AYAMLLTNKDTVPVRYGYDLYTDDGQYMATKSPYYDAITLLKARQYVAGGQSM----- 265

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 135.87 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-40

Perfect score: 7928

Sequence: 1 MENKHYKLHKVKKQWVTIA.....PGVYVFDKNGLAYPRVLN 1518

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_Main:*

- 1: /cgn2_6/prodata1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata1/pubpaa/US08_PUBCOMB.pep:*
- 3: /cgn2_6/prodata1/pubpaa/US09_PUBCOMB.pep:*
- 4: /cgn2_6/prodata1/pubpaa/US10A_PUBCOMB.pep:*
- 5: /cgn2_6/prodata1/pubpaa/US10B_PUBCOMB.pep:*
- 6: /cgn2_6/prodata1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7928	100.0	1518	4	US-10-383-930-40
2	7928	100.0	1518	5	US-10-797-821-40
3	3818	48.2	1554	4	US-10-383-930-38
4	3818	48.2	1554	5	US-10-797-821-38
5	3502.5	44.2	1430	3	US-09-740-274-6
6	3502.5	44.2	1430	4	US-10-383-930-36
7	3502.5	44.2	1430	5	US-10-797-821-36
8	3325.5	41.9	1590	4	US-10-383-930-37
9	3325.5	41.9	1590	5	US-10-797-821-37
10	3196.5	40.3	1475	3	US-09-740-274-2
11	3196.5	40.3	1475	4	US-10-383-930-34
12	3196.5	40.3	1475	5	US-10-797-821-34
13	3070	38.7	1375	3	US-09-740-274-4
14	3070	38.7	1375	4	US-10-383-930-35
15	3070	38.7	1375	5	US-10-797-821-35
16	2804	35.4	1365	4	US-10-383-930-39
17	2804	35.4	1365	5	US-10-797-821-39
18	2795.5	35.3	1497	5	US-10-484-218-18
19	2557.5	32.3	1595	5	US-10-484-218-20
20	2327	29.4	2057	4	US-10-417-280A-2
21	2142.5	27.0	1777	3	US-10-484-218-12
22	2093	26.4	1781	5	US-09-995-749A-2
23	2084	26.3	1006	5	US-10-484-218-22
24	2024.5	25.5	1771	5	US-10-484-218-14
25	1560.5	15.7	545	3	US-09-995-749A-10
26	1262	15.9	522	3	US-09-995-749A-11
27	1220.5	15.4	535	3	US-09-995-749A-13

28	1140.5	14.4	787	5	US-10-484-218-16	Sequence 16, Appl
29	1088	13.7	584	3	US-09-995-749A-12	Sequence 12, Appl
30	1055.5	13.3	525	5	US-10-484-218-23	Sequence 23, Appl
31	545.5	6.9	224	5	US-10-484-218-4	Sequence 4, Appl
32	541	6.8	223	5	US-10-484-218-10	Sequence 10, Appl
33	523	6.6	2710	4	US-10-011-366-6	Sequence 6, Appl
34	523	6.6	2710	4	US-10-354-774-6	Sequence 6, Appl
35	523	6.6	2710	4	US-10-271-012-6	Sequence 6, Appl
36	523	6.6	2710	4	US-10-729-123-6	Sequence 6, Appl
37	523	6.6	2710	4	US-10-729-039-6	Sequence 6, Appl
38	523	6.6	2710	5	US-10-729-527-6	Sequence 6, Appl
39	523	6.6	2710	5	US-10-727-898-6	Sequence 6, Appl
40	523	6.6	2710	5	US-10-728-696-6	Sequence 6, Appl
41	523	6.6	2710	6	US-11-001-241-6	Sequence 6, Appl
42	501	6.3	223	5	US-10-484-218-6	Sequence 6, Appl
43	447	5.6	811	4	US-10-011-366-7	Sequence 7, Appl
44	447	5.6	811	4	US-10-354-774-7	Sequence 7, Appl
45	447	5.6	811	4	US-10-271-012-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1

US-10-383-930-40
; Sequence 40, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-383-930-40

Query Match	100.0%;	Score 7928;	DB 4;	Length 1518;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1518;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MENKHYKLHKVKKQWVTIAVASVALATVLGGLSVTTSSVSADETQDKTIVTOSNGTTAS	60	
Db	1	MENKHYKLHKVKKQWVTIAVASVALATVLGGLSVTTSSVSADETQDKTIVTOSNGTTAS	60	
Qy	61	LVTSPATKEADKRTNTKEADVLTTPAKETNAVETATTTTQATAEAATTTATTADVAVAAV	120	
Db	61	LVTSPATKEADKRTNTKEADVLTTPAKETNAVETATTTTQATAEAATTTATTADVAVAAV	120	
Qy	121	PNKEAVTTDPAPVTTKEAEPATVKAEEVNVNTEVKAPEAAKDSVEAALSLKNIKNID	180	
Db	121	PNKEAVTTDPAPVTTKEAEPATVKAEEVNVNTEVKAPEAAKDSVEAALSLKNIKNID	180	
Qy	181	GKYYVNVNEDGSHKENFALTNGQLLYFGKDGALTSSSTYSFTPGTNTIVDGSINNAYD	240	
Db	181	GKYYVNVNEDGSHKENFALTNGQLLYFGKDGALTSSSTYSFTPGTNTIVDGSINNAYD	240	
Qy	241	SSEASFELIDGLTADSWYRPASIIKDGVTWQASTAEDFRPLLMAWPNVDQVNYLNYM	300	
Db	241	SSEASFELIDGLTADSWYRPASIIKDGVTWQASTAEDFRPLLMAWPNVDQVNYLNYM	300	
Qy	301	SKVFNLDKYSTDKQETLKVAAKDQIKIEQIKOAKSQTWLRETISAFVKTQPNWKE	360	
Db	301	SKVFNLDKYSTDKQETLKVAAKDQIKIEQIKOAKSQTWLRETISAFVKTQPNWKE	360	

Qy	361	TENYKGGGDEHLOGGALLYVNDSTRTPWANSYRRLNRATNTQTGTIDKSIILDESDPNH	420
Db	361	TENYKGGGDEHLOGGALLYVNDSTRTPWANSYRRLNRATNTQTGTIDKSIILDESDPNH	420
Qy	421	MGGFDFLLANDVDLSNPVQAQLNQIHYLMNWSIVMGDKDANPDGIRVDVNDVADM	480
Db	421	MGGFDFLLANDVDLSNPVQAQLNQIHYLMNWSIVMGDKDANPDGIRVDVNDVADM	480
Qy	481	LQLYTNYFXYGVNKESEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLPS	540
Db	481	LQLYTNYFXYGVNKESEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLPS	540
Qy	541	LAKPIKERTPAVSPLYNNTFNTQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNEKYGD	600
Db	541	LAKPIKERTPAVSPLYNNTFNTQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNEKYGD	600
Qy	601	ASGNVYFIRAHNNVDIIAEIIEKBEINPKSDGFTITDAEMKQAFIYNKMLSSDKYIT	660
Db	601	ASGNVYFIRAHNNVDIIAEIIEKBEINPKSDGFTITDAEMKQAFIYNKMLSSDKYIT	660
Qy	661	LNNIPAAVAMLQNMMETITRVYVYGDLYTDDGHYMETKSPYYDTIVNLMKSRIKYVSGQA	720
Db	661	LNNIPAAVAMLQNMMETITRVYVYGDLYTDDGHYMETKSPYYDTIVNLMKSRIKYVSGQA	720
Qy	721	QRSYMLPTDGKNDSDVELYRTNEVYTSVRYGKDINTANDTEGSKYSRTSGQVTLVANNP	780
Db	721	QRSYMLPTDGKNDSDVELYRTNEVYTSVRYGKDINTANDTEGSKYSRTSGQVTLVANNP	780
Qy	781	KLNDQSAKLNVEMGKIHANQYRALIVGTADGINKFTSDADAIAAGYVYKETSNGVLTF	840
Db	781	KLNDQSAKLNVEMGKIHANQYRALIVGTADGINKFTSDADAIAAGYVYKETSNGVLTF	840
Qy	841	GANDIKGYETFDMSGFVAVVMPVPGASDNQDIIRVAPSTEAKGEGELTLKATEAYDSQLIYE	900
Db	841	GANDIKGYETFDMSGFVAVVMPVPGASDNQDIIRVAPSTEAKGEGELTLKATEAYDSQLIYE	900
Qy	901	GFSNFQTIIPDGDSPSYVYTRKIAENVDLFLKSGWGVTSFEMAPQFVSADDDGTFILDSVIQNGY	960
Db	901	GFSNFQTIIPDGDSPSYVYTRKIAENVDLFLKSGWGVTSFEMAPQFVSADDDGTFILDSVIQNGY	960
Qy	961	AFADRYDLAMSKNNKYGSKEDLRDALKALHKAGIOAIADWVPDQIYQLPCKEWTATRTD	1020
Db	961	AFADRYDLAMSKNNKYGSKEDLRDALKALHKAGIOAIADWVPDQIYQLPCKEWTATRTD	1020
Qy	1021	GAGRKIADAIIDHSLYVANSKSGKDYOAKYGEEFLAELKAKYPEMFKVNMISTGKPIDD	1080
Db	1021	GAGRKIADAIIDHSLYVANSKSGKDYOAKYGEEFLAELKAKYPEMFKVNMISTGKPIDD	1080
Qy	1081	SVKLLQWKAEYFNGTNVLERGVGYVLSDEATGKYFTVTKEGNFIPLOLTGKEKVIITGFS	1140
Db	1081	SVKLLQWKAEYFNGTNVLERGVGYVLSDEATGKYFTVTKEGNFIPLOLTGKEKVIITGFS	1140
Qy	1141	DGKGITYFCTSGTQAASAFVTFNGNTYYFDARGHMVNTSEYSPNGKDVTYRFLPNGIMLSN	1200
Db	1141	DGKGITYFCTSGTQAASAFVTFNGNTYYFDARGHMVNTSEYSPNGKDVTYRFLPNGIMLSN	1200
Qy	1201	AFYIDANGNTYLYNSKGQMYKGGYTKFDVSETDKDGKESKVVKFRYFTNEGVMAGKVTVI	1260
Db	1201	AFYIDANGNTYLYNSKGQMYKGGYTKFDVSETDKDGKESKVVKFRYFTNEGVMAGKVTVI	1260
Qy	1261	DGFTQYFGBDGFQAKDLVTFKGKTYTFDAHTGNGIKOTWRNINGKWIYFDANGVAATGA	1320
Db	1261	DGFTQYFGBDGFQAKDLVTFKGKTYTFDAHTGNGIKOTWRNINGKWIYFDANGVAATGA	1320
Qy	1321	QVINGOKLYFNEDEGSQVKGWKNADGTSKYKEGFGELVTNEFFTTTDCGNVWYAGANGK	1380
Db	1321	QVINGOKLYFNEDEGSQVKGWKNADGTSKYKEGFGELVTNEFFTTTDCGNVWYAGANGK	1380
Qy	1381	TVTGAQVINGQHLVFNADGSQVKGWVKNADGTYSKYNASTGERLTNEFFTTGDNWNYI	1440
Db	1381	TVTGAQVINGQHLVFNADGSQVKGWVKNADGTYSKYNASTGERLTNEFFTTGDNWNYI	1440

Qy	1441	GANGKSVTGEVKIGDDTYFFAKDGKQVKGQTVSAGNGRISYYYGDSGKRAVSTWIEIQPG	1500
Db	1441	GANGKSVTGEVKIGDDTYFFAKDGKQVKGQTVSAGNGRISYYYGDSGKRAVSTWIEIQPG	1500
Qy	1501	VYVYFDKNGLAYPPRVLN	1518
Db	1501	VYVYFDKNGLAYPPRVLN	1518
RESULT 2			
US-10-797-821-40			
; Sequence 40, Application US/10797821			
; Publication No. US20050031633A1			
; GENERAL INFORMATION:			
; APPLICANT: Smith, Daniel J.			
; APPLICANT: Taubman, Martin A.			
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens			
; FILE REFERENCE: 25669-020			
; CURRENT APPLICATION NUMBER: US/10/797,821			
; CURRENT FILING DATE: 2004-03-09			
; PRIOR APPLICATION NUMBER: 10/383,930			
; PRIOR FILING DATE: 2003-03-07			
; PRIOR APPLICATION NUMBER: 60/363,209			
; PRIOR FILING DATE: 2002-03-07			
; PRIOR APPLICATION NUMBER: 60/402,483			
; PRIOR FILING DATE: 2002-08-08			
; PRIOR APPLICATION NUMBER: 09/290,049			
; PRIOR FILING DATE: 1999-04-12			
; PRIOR APPLICATION NUMBER: 60/081,550			
; PRIOR FILING DATE: 1998-04-13			
; PRIOR APPLICATION NUMBER: 60/115,142			
; PRIOR FILING DATE: 1999-01-08			
; NUMBER OF SEQ ID NOS: 45			
; SOFTWARE: Patent in version 3.2			
; SEQ ID NO 40			
; LENGTH: 1518			
; TYPE: PRT			
; ORGANISM: Streptococcus salivarius			
US-10-797-821-40			
Query Match 100.0%; Score 7928; DB 5; Length 1518;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MENKHYKLHKVKKQWVTIAVASVALATVILGGLSVTTSSVSADETQDKTVTQSNSTTAS	60
Db	1	MENKHYKLHKVKKQWVTIAVASVALATVILGGLSVTTSSVSADETQDKTVTQSNSTTAS	60
Qy	61	LVTSPKATKADKRTNTEADVLTPAKETNAVETATTTNTQATAEAATTTATTADVAVAV	120
Db	61	LVTSPKATKADKRTNTEADVLTPAKETNAVETATTTNTQATAEAATTTATTADVAVAV	120
Qy	121	PNKEAVVTTDPAVTTTEKAEQPAVTKAEVNVNTEVKAPPAALKDSEVEAALSILKNIKNID	180
Db	121	PNKEAVVTTDPAVTTTEKAEQPAVTKAEVNVNTEVKAPPAALKDSEVEAALSILKNIKNID	180
Qy	181	GKYVYVNDGSGHENFAITVNGQLYFGKDGALTSSSTYSFTPGTTNIVDGSINNRAVD	240
Db	181	GKYVYVNDGSGHENFAITVNGQLYFGKDGALTSSSTYSFTPGTTNIVDGSINNRAVD	240
Qy	241	SSEASFELIDGVLTDADSVYRPASIIKQVVTWQASTAEDFRPLLMAWPNVDTOVNYLNYM	300
Db	241	SSEASFELIDGVLTDADSVYRPASIIKQVVTWQASTAEDFRPLLMAWPNVDTOVNYLNYM	300
Qy	301	SKVFNLDAKYSSTDQETLKVAAKDIQIKIEQIKQAEKSTQWLRETIISAFVKTQPOWNKE	360
Db	301	SKVFNLDAKYSSTDQETLKVAAKDIQIKIEQIKQAEKSTQWLRETIISAFVKTQPOWNKE	360
Qy	361	TENYKGGGDEHLOGGALLYVNDSTRTPWANSYRRLNRATNTQTGTIDKSIILDEQSDPNH	420
Db	361	TENYKGGGDEHLOGGALLYVNDSTRTPWANSYRRLNRATNTQTGTIDKSIILDEQSDPNH	420
Qy	421	MGGFDFLLANDVDLSNPVQAQLNQIHYLMNWSIVMGDKDANPDGIRVDVNDVADM	480

Db 421 MGGDFLLANDVLSNPVQAEQLNQIHVLMNWSIVMGDKDANFGIRVDVNDYDAD 480
QY 481 LQLYTYNFRYYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLFS 540
Db 481 LQLYTYNFRYYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLALLFS 540
QY 541 LAKPIKERTPAVSPLYNNTFTTQREKTDWINKOGSKAYNEDGTQKSTIGKYNEKYGD 600
Db 541 LAKPIKERTPAVSPLYNNTFTTQREKTDWINKOGSKAYNEDGTQKSTIGKYNEKYGD 600
QY 601 ASGNYVFIKADHNNVDIIABIIKKEINPKSDGFTTDAEMKQAFIYNKMLSSDKKYYT 660
Db 601 ASGNYVFIKADHNNVDIIABIIKKEINPKSDGFTTDAEMKQAFIYNKMLSSDKKYYT 660
QY 661 LNNIPAAYAVMLQNMTITRYYGDLTYDDGHYMETKSPYYDTIVNLMKSRIKYVSGGQA 720
Db 661 LNNIPAAYAVMLQNMTITRYYGDLTYDDGHYMETKSPYYDTIVNLMKSRIKYVSGGQA 720
QY 721 QRSYWLPTDGMNDSDVELYRTNEVYTSVRYGKDIINTANDTEGSKYSRTSGQVTLVANRP 780
Db 721 QRSYWLPTDGMNDSDVELYRTNEVYTSVRYGKDIINTANDTEGSKYSRTSGQVTLVANRP 780
QY 781 KNLNDOSAKLVNEMGKIHANQYRALIVGTADGINKFTSDADAIAAGYVKETDSNGVLTF 840
Db 781 KNLNDOSAKLVNEMGKIHANQYRALIVGTADGINKFTSDADAIAAGYVKETDSNGVLTF 840
QY 841 GANDIKGYETFDMSGFVAWVVPVPGASDNQDIRVAPSTEAKKEGELTLKATEAYDSQLIYE 900
Db 841 GANDIKGYETFDMSGFVAWVVPVPGASDNQDIRVAPSTEAKKEGELTLKATEAYDSQLIYE 900
QY 901 GFSNFOTIPDGDSPSVYNNRKAENVDLFSKMGVTSFEMAPQVSDGDTFLDSVIQNGY 960
Db 901 GFSNFOTIPDGDSPSVYNNRKAENVDLFSKMGVTSFEMAPQVSDGDTFLDSVIQNGY 960
QY 961 AFADRYDLAWSKNNKYGSKEDLDALKAHKGIAQIADWPDQIYQLPGKEVVTATRD 1020
Db 961 AFADRYDLAWSKNNKYGSKEDLDALKAHKGIAQIADWPDQIYQLPGKEVVTATRD 1020
QY 1021 GAGRIADIADHSLVYANSSKSGDKYQAKYGEFELAEKAKYPEMFKVNMISTGKPIDD 1080
Db 1021 GAGRIADIADHSLVYANSSKSGDKYQAKYGEFELAEKAKYPEMFKVNMISTGKPIDD 1080
QY 1081 SVKLAKQWKAERYFNGTNVLRGVGYVLSDBATGKYFTVTKEGNFIPLQLTGKEKVIITGFS 1140
Db 1081 SVKLAKQWKAERYFNGTNVLRGVGYVLSDBATGKYFTVTKEGNFIPLQLTGKEKVIITGFS 1140
QY 1141 DGKGITYFCTSGTOAKSAFVTNGNTYYFDARGHMVTNSEYSPNGKDVTRFPLNGIMLSN 1200
Db 1141 DGKGITYFCTSGTOAKSAFVTNGNTYYFDARGHMVTNSEYSPNGKDVTRFPLNGIMLSN 1200
QY 1201 AFYIDANGNTYLYNSKGOMYKGYTKFDVSETDKDGESKVVVKFRYFTNEGVMAGVTVI 1260
Db 1201 AFYIDANGNTYLYNSKGOMYKGYTKFDVSETDKDGESKVVVKFRYFTNEGVMAGVTVI 1260
QY 1261 DGFTQYFGEDGFQAKDLVTFKGKTYFFDAHTGNGIKDTRWNINGKYYFDANGVAATGA 1320
Db 1261 DGFTQYFGEDGFQAKDLVTFKGKTYFFDAHTGNGIKDTRWNINGKYYFDANGVAATGA 1320
QY 1321 QVINGOKLYFNEDGSGQGVGVKNADGTYSKYKEGFGELVITNEFFTTDGNVWYAGANGK 1380
Db 1321 QVINGOKLYFNEDGSGQGVGVKNADGTYSKYKEGFGELVITNEFFTTDGNVWYAGANGK 1380
QY 1381 TVTGAQVINGOHLVFNADGSGQGVGVKNADGTYSKYNASTGBRLTNEFFTTGDNWYI 1440
Db 1381 TVTGAQVINGOHLVFNADGSGQGVGVKNADGTYSKYNASTGBRLTNEFFTTGDNWYI 1440
QY 1441 GANGKSVTGEVKIGDDTYFFAKDGQVKQGTYSAGNGRISYYYGSGKAVSTWIEIQPG 1500
Db 1441 GANGKSVTGEVKIGDDTYFFAKDGQVKQGTYSAGNGRISYYYGSGKAVSTWIEIQPG 1500
QY 1501 VVYVFDKNGGLAYPPRVLN 1518

Db 1501 VVYVFDKNGGLAYPPRVLN 1518
RESULT 3
US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38
Query Match 48.2%; Score 3818; DB 4; Length 1554;
Best Local Similarity 50.1%; Pred. No. 1.2e-198;
Matches 793; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
QY 1 MENKIHYKLHKVKQWVTIIVASVALATVGLGLSVTSSVSADETDQDKTVTQSNSTTAS 60
Db 1 MEKLLHYKLHKVKQWVTIIVASIGLVSLVG----AGTVSA---EDKVANDTTAQATVG 52
QY 61 LVTSPEATKEADKRTTKTKEADVLTAKETNAVETAT-TTNTQATATAAATTAATADVA-VA 118
Db 53 VDTGQDQATNTDANTNTDTDADQSANNTNQAGSDQSNQDQAKQDANTANDRNQADNS 112
QY 119 AVPNKEAVVTTDAPAVTTEKAEQPAVKAEVNTVEKPAEALKDSEVEAALSLEKNIKN 178
Db 113 QTDNQATQDQATSPATDGTSVORRDA--ANVATAADQEGQTAPSEQEKSAALSLENVKL 169
QY 179 IDGKYVYVNEGDGSHKENFAITVNGQLLYFGKD-GALTSSSTYSFTPTGTTNIVDGFSSINR 237
Db 170 IDGKYVYVQADGSGYKKNFAITVNGQMLYFDSDTGALSTSTYSFSQGTTLNLVDDFRSHNK 229
QY 238 AYDSEASFEILDGVLTDADSWYRPASIIKDGVTWQASTABDPRPLLMANWPNVDVTQVYL 297
Db 230 AYDSTAKSFELVNGYLTANSWYRPAGILRNGQTWEASNENDLRPVLMSWPKDQTVAYV 289
QY 298 NYMSKVFNLD-KAKYSSSTDKQETLKVAADQIQIKIQIOAEKSTOWLRETISAFVKTOPQ 356
Db 290 NTMNYKLSANEVETNETSVDNLNKEAQSIQTKIQQKITSNDNSTOWLRTAMEAFVAAQPK 349
QY 357 WKNETENYSKGGEDHLOGGALLYVNDSTRTPWANSDYRLNRTATNTQGTIDKSIILDEOS 416
Db 350 WNMSTENFKG--DHLQGGALLYTNSDLTPWANSDYRLNRTPTQDGT--KKYFTEGG 404
QY 417 DPNHMGGFDFLLANDVDLSNPVQAEQLNQIHVLMNWSIVMGDKDANFGIRVDVNDYDAD 476
Db 405 E---GGYEFLLSNDVDNSNPVQAEQLNQHLNMGWDIVMGDKDANFGIRVDVNDYDAD 460
QY 477 DADMLQVLTNTYFREYGVNKSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLA 536
Db 461 NADLLQVTSNFPKONYKVTDSSEANALAHISVLEAWSLNDNHNDKTDGAALAMENKORLA 520
QY 537 LFLSLAKPIKERTPAVSPLYNNTFTTQREKTDWINKOGSKAYNEDGTQKSTIGKYNE 596
Db 521 SLAVLTK-----QPQQRIDLSNLSISESVNKERAND-----T 551
QY 597 KYGDASGNYVFIKADHNNVDIIABIIKKEINPKSDGFTTDAEMKQAFIYNKMLSSD 656
Db 552 AYGDITPTYSFVRAHDSVQTVIAKIVKEKIDTNSDGYTFTLDQLKDAFKIYNEDMAKVN 611

```
QY 657 KYITLNNIPAAVAVMLQNMETITRVYVYGLYTDGHHYMETKSPYVDTIVNLNKSRIKYVS 716
Db 612 KYITHYNIIPAAVALLSNMESPRVYVYGLYTDGQYMAKSPYYDAIATMLQGRYAYVS 671
QY 717 GGOAQRSYWLPDTGKMDNSDVELYRTNEVYTSVRVYKDIWMTANDTEGSKYSRTSGQVTLV 776
Db 672 GQOSEVH-----KNGNQLSSVRVYGDILMSADDTQGTDLSTRSGLVTLV 718
QY 777 ANPKLNLQDSAKLANVEMGKIHANQKRALIVGTADGINKFTSDADAIAAGVYKETSQ 836
Db 719 SNDPNLDLGDS-LTVNMGRAHANQAYRPLILGTGQVQSLKSD--TNIVKYTDANG 774
QY 837 VITFGANDIKGVETFDMSGFVAWVPVGASDNQDIRVAPSTEAKKEGELTLKATEYDQ 896
Db 775 NLTFADDIKGYSTVDMGYLAWVPVGAKDQDVRVAADTNQKADGK-SLKTSAAALDQ 833
QY 897 LIYEGFSNFQTTIDGSDPSVYNNRKTAEVNDLPKSGVTSFEMAPQVFSADGTFLDSVI 956
Db 834 VIYEGFSNFQDF--ANNADYTNKKIAENADFKKLGITTSFEMAPQVVSATDGSFLDSII 891
QY 957 QNGYAFADRYDLAMSKNNKYSGKEDLRDALKALHKAGIOAIADWVPDQIYQLPGKEVVT 1016
Db 892 QNGYAFSDRYDLAMSKNNKYSGKODLANALKALHANGIOAIADWVPDQIYQLPGEEVVT 951
QY 1017 TRTDGAGRIADAIDHLSYVANSKSGDYQAKYGGEFLELAKAYPMPFKNMISTGK 1076
Db 952 KRTNSYGNPTFDAYINNALYATNTKSSGSDYQAGYGAFLDELKAKYPMFTVMNISTGK 1011
QY 1077 PIDDSVKLQWKAERYFNGTNVLERGVYVLSDEATGKYFTVTKEGNFIPLOLTGKEKVI 1136
Db 1012 PIDPSTKIKQWKAERYFNGTNVLERGVYVLSDDATGKYFTVNEGDFLPASTFGQNAKT 1071
QY 1137 GFSSDGKGYTYFGTSGTQAKSAFVTFNGNTYTFDARGHMTVNSEYSPNGKDVYRFLPNCI 1196
Db 1072 GFYVDTGWAYYSTSGNAKAVNSFIYEGHYVYFDKDHMTVSGYKAEEDNDY-FLPNCI 1130
QY 1197 MLSNAPYIDANGNTYLYNSKGOMYKGG--YTFPDSVETDKGKESKVKVFRFTNEGVNA 1254
Db 1131 QMRDAIYODAOGNSYYVYGRIGLYKGDNWPYFVDPNNANK-----TVFRYFDANNVNA 1183
QY 1255 KGVTVIDGFTQYFGEDGFOAKDLVTFKGTYYFDAHTGNGIKDWRNTNGKWIYFDANG 1314
Db 1184 IGRNMYGOTYTFDENGFOAKGOLLTDKGTHTYFDEDNGAMAKFPVNVGDDWYMDNG 1243
QY 1315 VAATGAQVINGOKLYFN-EDGSOVKGVVYKNADGTYSKYKEGFGELVTINEFFTTDGNVY 1373
Db 1244 NAVKGQYPVNNQILYFNPETGVQVKGQFITDAQRTSYDYDANGSALKSSGFFTPNGSDWY 1303
QY 1374 YA-----GANGKTVTGAQVINGOHLYFNAD-GSOVKGVVKN 1409
Db 1304 YAEANGVYVYKFGQVAENQDQWYVFDQTTGQAKGAKVDDGRDLVFNPDGSGVQVKGDFATD 1363
QY 1410 ADGTSYKNASTGERLTNTEFFTTGDNWYVYI GANGSVTVGEVKI----- 1453
Db 1364 ESGNTSFYHGDNGDKVVGFFFTGNNAWYIADNNGNLVKGFEIDGKWKTHFDEVTCQQA 1423
QY 1454 -----GDDTYFFAKDGKQVKGQVTSAGNRSYVYVYDGSCKEAVSTWIBIQPGVYVYFDK 1507
Db 1424 GAALVNGQQLYFDVDSGIVQVKGDFVTDGQNTSYVDNSGDKKXNGVFFTTGDNWYVADG 1483
QY 1508 NG 1509
Db 1484 QG 1485
```

RESULT 4

```
US-10-797-821-38
; Sequence 38, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
```

```
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-38
```

Query Match

```
Best Local Similarity 48.2%; Score 3818; DB 5; Length 1554;
Matches 783; Conservative 215; Mismatches 434; Indels 130; Gaps 24;
```

```
QY 1 MENKHYKHLKHKVQWVTIAVASVALATVLGLSVTSVSADETQDKTIVTQSNSTTAS 60
Db 1 MEKHLHYKHLKHKVHWVTIAVASIGLSLVG-----AGTVSA--EDKVANDTTAQATVG 52
QY 61 LVTSPEATKEADKRTNITKEADVLTTPAKETNAVETAT-TNNTQATAEAAATATTADVA-VA 118
Db 53 VDTGQDQATNTDANTNTTDTADQASANTNQDQAGSDQSNNOQAKQDQTANTDRNQADS 112
QY 119 AVPNKEAVVTTDAPAVTTEKABEQPATVKAEEVNTVEKPAEALKDSEVEAALSLKNIKN 178
Db 113 QPTDNOATDQATSPATDGTSGVQRDA--ANVATAADQEGQTPAPSEKSAALSUDNVKL 169
QY 179 IDGKYVYVNEGDSGHKENFAITVNGQLLYFGKD-GALTSSSTYSFTPGTNNIVDGFINN 237
Db 170 IDGKYVYVQADSGYKKNFAITVNGQMLYFSDTGLSSTSTYSFSGQTTNLVDFFSSH 229
QY 238 AYDSEASFEILDGVLTDASWYRPAIIKDGVTWQASTAEADPRPLLMANWPNVDTOVYL 297
Db 230 AYDSTAKSFELVNGVLTANSWYRPAIILRNGQTWEASNENDLRPVLMSWPKDQTQVAYV 289
QY 298 NYMSKVFNL-DAKYSTDQKQETLKYAAKDOIQIKIQKIOAEKSTOWLRETISAFVKTPQ 356
Db 290 NYNMKYL SANETEVNTS QVDLNKEAOSIQTKIBOKITSDNSTOWLRTAMEAFVAAQPK 349
QY 357 WNKETENTYSGGGEHLOGGALLYVNDSTRTPWANSDYRRLNRTATNQTGTIDKSLIDEOS 416
Db 350 WNNSTENFKG--DHLGGALLYTNSDLTPWANSDYRLNRTPTQODGT--KKYFTEGG 404
QY 417 DPNHNGGDFLLANDVDLSNPVVQAEQLNQHLYLNMWGSIVMGDKDANFDGRVADVNV 476
Db 405 E---GGVEFLLSNDVDNSNPVVQAEQLNQLHYLNMWGDIVMGDKDANFDGRVADVNV 460
QY 477 DADMLQLYTNPREYGVNKSSEANALAHISLVEASLNDNHNNDKTDGAALAMENKORLA 536
Db 461 NADLLQVYSNYFKDNYKVTDSEANALAHISILEASLNDNQNDNEDTNGTALSIDNSRLT 520
QY 537 LLFSLAKPIKERTPAVSPLYNNTFNTRQDEKTDWINKDGSKAYNEDGTVKQSTIGKYNE 596
Db 521 SLAVLTK-----QPGQIDLSNLISESVNKERAND-----T 551
QY 597 KYGDASGNVFTIRAHNNVQDIIABIIKKKINPKSGDFTITDAEMKQAFPIYKNMLSSD 656
Db 552 AYGDPTIPTYSFVRAHDSEVQTVIAKIVKEKIDTNSDGYTFTLDQLDAFKIYNEDMAKNV 611
QY 657 KKYTLNNIPAAVAVMLQNMETITRVYVYGLYTDGHHYMETKSPYVDTIVNLNKSRIKYVS 716
```



```
Db 613 LKQAFKIYNEDMRQAKKKTQSNIPAYALMLSNKDSITRLYYGDMYSDGQYMATKSPY 672
Qy 701 YDTIVNLMSRKIKYVSGQAQRSYMLPTDGKMDNSDELRYRTNEVYTSVRYGKDINTAND 760
Db 673 YDAIDTLLKARIKYAAGGQDMKITVY--EGDKSHMD---WDYTVGLTSVRYGTGANEATD 727
Qy 761 TEGSKYSRSTGGQVTLVANNPKLNDQSAKLNVEMGKIHANOKYRALLIVGTADGKINFISD 820
Db 728 -QGSEATKTQGMVITSSNNPSLKLNDQKVIYVMGAAHKNQYRPELLTLLTKDGLTSYTS- 785
Qy 821 ADAIAAGYVYKETDNGVLTFGANDIKGYETFDMSGFVAVMVPGASDNDQIRVAPSTEAK 880
Db 786 -DAAKSLYRKNDKGELVFDASDIQGYLNPQVSGYLAVMVPGASDNDQVRVAASNKAN 844
Qy 881 KEGELTLKATEAYDSQLIYEGFSNFOT-IPDGSDPSSVYTRNRKIAENVDLFKSWGVTSPFM 939
Db 845 ATGOV-YESSALSALQLIYEGFSNFQDFVTKDS- -YTNKKIAQNVLFKSWGVTSPFM 900
Qy 940 APQFVSADGTFLDSVIQNGYAFADRYDLAMSKNNKYGSKEDLRDALKAHLKAGIQAIAD 999
Db 901 APQYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQDMINAVAKALHKSIGQVIAD 960
Qy 1000 WYPDQIYLPQGEVVTATRTDAGRKIADAIIDHSLYVANSKSGKDYOAKYGGEPFASL 1059
Db 961 WYPDQIYLPQGEVVTATRVNDYGEYRKDSIKNTLYAANTKSNKDYQAKYGGEPFASL 1020
Qy 1060 KAKYPMFKNMISTGPKIDDSYKLLKQWKAEPNGTIVLERGVYGVYLSDEATGKPYFTVK 1119
Db 1021 AAKYPSIFNRTQISNGCKIDPSEKITAWKAKYFNGTNIILGRGVYVLDKONASDKYFELKG 1080
Qy 1120 EGNFIPLOUTGKEKVTITGFSDDGKGTFTGSGTQAKSAFV-TFNGNTYTFDARGHMTN 1178
Db 1081 NOTYLPKQMTNKE-ASTGTFVNDGNGMTFTYSTGYQAKNSFVQDAKGNWYTFDNGHMYG 1139
Qy 1179 SEVSPNGKDYVFLPGLNGIMLSNIFYIDANGNTVLYNSKGOMYKGGYTFKFDVSETDKDGE 1238
Db 1140 LQ-QLNG-EVQFSLSGVQLRESFLENADGSKNYFGLGNRYSGNYSFNDNS----- 1190
Qy 1239 SKVYFRYFTNEGVMKAVGVTVIDGFTQYFCEGDFQAKDLVLT-FKGKTYTFDAHTNGIK 1297
Db 1191 ----KWRYFDASGVMAVGLKTINGNTQYFDQGYQVKGAWITGSDGKKRYFDDGSGNMAV 1246
Qy 1298 DTRWN-INGKWTYFDANGVAATGAQVINGOKLYFNEDGSOVKGVVKNADGTYSKYKEGF 1356
Db 1247 NRPANDKNGDWYLYNSDGIALGVQTINGKTYTFGQDGKQIKKIITD-NGKLYFLANS 1305
Qy 1357 GELVTNEFTTDGNVWYVYAGANGKVTGAQVINGOHLVFNADGSOVKGGVKNADGTYSK 1416
Db 1306 GELARNIFATDSQNNWYIFGSDGVAVTGSGTTIAGKKLYFASDGKQVKGFSV-TYNGKVHY 1364
Qy 1417 YNASTGERLTNEFTTGDNNWYVYIGANGKSVTGEVKIGDDTYFFAKDGRQVKGQTYSAGN 1476
Db 1365 YHADSGELQVNRFEADKGNWYLYDSNGEALTGSGQRINDQVFFTRGKQVKGD-VAYDE 1423
Qy 1477 GRISY 1482
Db 1424 RRLVY 1429
```

RESULT 6

```
US-10-383-930-36
; Sequence 36, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
```

```
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-36
```

```
Query Match 44.2%; Score 3502.5; DB 4; Length 1430;
Best Local Similarity 48.1%; Pred. No. 1.5e-181;
Matches 724; Conservative 247; Mismatches 434; Indels 101; Gaps 32;
```

```
Qy 1 MENKHYKHLKVKQWVTIAVASVALATVLGGLSVTTSVSAD-----ETQDKVTVO-SNGS 56
Db 1 METKERYVMHKVGHVTVAVAS-GLITL--GTTTLGSSVSAETBQOTSDKVVTQKSEDD 57
Qy 57 TTASLAVTSPEATKKEADKRTNTKEADVLTPAKETNAVETATTTNTQATAEAAAT-ATTADV 115
Db 58 KAASSTQTDAPKTQQAQTEQTAO-----SQANVADTSTSTKETPSONIITQANSDDK 112
Qy 116 AVAAVPEKNAVVTDDAPAVTTEKAEQEPATKVAEVVNTVEKA-----PEAALKDSEVBA 169
Db 113 TVTNTKSEAAQTSSE-----RTKQSEEAQTASSQAL-TQAKAELTKQRTAAQENKNPVD 167
Qy 170 ALSLKNINIDKYYVNVNEDGSHKENFALTVMGQLLYFGKD-GALTSSSTYSFTPGTTNI 228
Db 168 LAAIPNVKQIDGKYIYIGSDGQPKNFALTVMNNKVLVFDKNTGALTDTTSQYQFKGLTKL 227
Qy 229 VDGFSINNRAYDSSEASPELIDVLTADSWYRPASI IKDGVTVQASTAEDFRPLLMAWWP 288
Db 228 NNDYTPHQIVNFENTSLETIDNYVTADSWYRPKDLKNGKWTWASSSDLPRLMSWNP 287
Qy 289 NYDVTQVNLVNMYSKY-FNLDAKYSTDKQETLKVAAKDOIQIKIEQIKQAEKSTQWLRETI 347
Db 288 DKQQTQIAYLVNMVQOGLTGENTYADSSQESLNLAAQTQVQVKIETKISQTOQTQMLRDI 347
Qy 348 SAFVTKTQPNKETSNYKGGEDHLOGGALLYVNDSTPTPMANSYRRLNRTATNQTGHI 407
Db 348 NSFVKTQPNMNSQTSDETSAGEKHLQGGALLYSNSDKTAYANSDYRLNRTPTSTQTK- 406
Qy 408 DKSILDEQSDPNHMGDFDILLANDVLSNPVVAQSLNOIHYLMNWSGTVMGDKDANFDG 467
Db 407 -----PKYFEDNSSGGYDFLLANDIDNSNPVVAQSLNMLHYLMNWSGTVMGDKDANFDG 461
Qy 468 IRVDAVDNVDADMLQLYTNYFREYTGYNKSEANALAHISVLSEAWSLNDNHNDKTDGAAL 527
Db 462 VRVDAVDNVDNADLLQIASDYLKAHYGVDSKSEKNAIHLISLEAWSNDNDPQYNKDTKGAQL 521
Qy 528 AMENKORLALLPSLAKPIK-----ERTPAVSPLYNNTFTNTTQORDEKTDWINKGSKAY 580
Db 522 PIDNKLRLSLLYALTRPLEKADSNKNEIRSGLEPVITNSLN----- 562
Qy 581 NEDGTVKOSTICKYNEKYGDASGNVYFIRAHNNVDIIAEIIKKEINPKSGFTITDAE 640
Db 563 -----NRSAEKGNSEK-----MANIFIRAHUSEVQTVJAKIIKAQINPKTDLGLTFLDE 612
Qy 641 MKQAFETYNKMDLSSDKKYLNNIPAAVAVMLQNNMETITRVYVYDLYTDDGHYMETKSPY 700
Db 613 LKQAFKIYNEDMRQAKKKTQSNIPAYALMLSNKDSITRLYYGDMYSDGQYMATKSPY 672
Qy 701 YDTIVNLMSRKIKYVSGQAQRSYMLPTDGKMDNSDELRYRTNEVYTSVRYGKDINTAND 760
Db 673 YDAIDTLLKARIKYAAGGQDMKITVY--EGDKSHMD---WDYTVGLTSVRYGTGANEATD 727
Qy 761 TEGSKYSRSTGGQVTLVANNPKLNDQSAKLNVEMGKIHANOKYRALLIVGTADGKINFISD 820
Db 728 -QGSEATKTQGMVITSSNNPSLKLNDQKVIYVMGAAHKNQYRPELLTLLTKDGLTSYTS- 785
Qy 821 ADAIAAGYVYKETDNGVLTFGANDIKGYETFDMSGFVAVMVPGASDNDQIRVAPSTEAK 880
Db 786 -DAAKSLYRKNDKGELVFDASDIQGYLNPQVSGYLAVMVPGASDNDQVRVAASNKAN 844
```



```
Db 901 APQYSSSEDSFLDSIIQNGYAFEDRYDLAMSKNNKYSGQDMINAVKALHKSIGQVIAD 960
Qy 1000 WVPDQIYQIPGKEVVTATFTDAGAKRIADAIIDHSLYVANSKSKGYQAKYGGSEFLABEL 1059
Db 961 WVPDQIYQIPGKEVVTATRVNDYGEYKDSIKNTLYAANTKSGKDYQAKYGGAFSEL 1020
Qy 1060 KAKYEMFKVMNISTGKPIDDSVKLGKQWKAERYFNGCTNVLERGVYVLSDEATGKYFTVTK 1119
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAUKAKYFNGNIIIGRGVGVYVLDKNASDKYFELKG 1080
Qy 1120 EGNFIPLOLTGKEKVTITGSSDGKITTYFGTSGTQAKSAFV-TFNGNTYFFDARGHMYTN 1178
Db 1081 NQTYLPKQNTKE-ASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFFDNGHMYVG 1139
Qy 1179 SEYSPNGKDYVRLPLNGIMLSNAFYIDANGNTYLYNSKQMYKGYTKPDPVSETDKDGKE 1238
Db 1140 LQ-QLNG-EVQYFSLNGVQLRESFLENADGSKNYFCHLGNRYSGYYSFDNDS----- 1190
Qy 1239 SKVVKFRYFTNGVMAKGVTVIDGFTQYFEGEDGFOAKDKLVT-FKGKTYVFDHTGNGIK 1297
Db 1191 ----KWRYFDASGVMAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKKRYFDDGSGNMAV 1246
Qy 1298 DTRWN-INGKWKYVFDANGVAAAGAOVINGOKLYFNEGDSQVKGGVVKNADGYSKYKEGF 1356
Db 1247 NPFANDKGDWYLYNSDGLALVGVTINGKTYFFGQDGKQIKGIITD-NGKLYFLANS 1305
Qy 1357 GELVTNEFTTDGNWYVYAGAKVTGGAQVINGOHLYFNADGSOVKGGVKNADGYTSK 1416
Db 1306 GELARNIFATDSQNNWYFGSDGVAVTGQTIAGKKLYFASDGKQVKSFV-TYNGKVHY 1364
Qy 1417 YNASTGERLTNEFTTGDNNWYVYIGANGSKVTEGVKIGDDTYFFAKDGKGVKQTVSAGN 1476
Db 1365 YHADSGELQVNRFEADKQGNWYVYLDNSNGEALTQSGQRINDQRVFFTREKGQVKGQ-VAYDE 1423
Qy 1477 GRISY 1482
Db 1424 RLLVY 1429

RESULT 8
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Taubman, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 37
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match 41.9%; Score 3325.5; DB 4; Length 1590;
Best Local Similarity 45.0%; Pred. NO. 7.2e-172;
Matches 710; Conservative 239; Mismatches 456; Indels 173; Gaps 33;

Qy 1 MENKHYKHLKVKKQWVTIYAVASVA-LATVLGSLVTTSSVSADETDKTVTQNSGTTA 59
Db 1 MEKNVRFKWKVKRWVTVLSVASATMLASALG-----ASVASAD-----TDTASDD----- 46
Qy 60 SLVTSPEATKEADKRNTEADVLTPAKETNAVETATTTTQATABAAATATTADVAVAA 119
Db 47 ----SNQAVVYTGDTTNNQATD-----QTSIAATAT-----SEQSASTDAATDAQSAA 90
```

```
Qy 120 VPKAEAVVTTDAPVTTTEKAEQEPATVKAEEVNV-----TEVKAPEAALKDSEVEAL 171
Db 91 EOTQGTASTTAAQTQTTTNAME-AKWVPTENENOGFTDEMLAEAKNVATABSIDSIPSLIA 149
Qy 172 SLKIKNIDGKYVYVNEGSHKENFAITVNGQLLYFGDKGALTSSTSYSTFTPTGTTNIVDG 231
Db 150 KMSNVKQVDGKYVYDQDGNVKKFPAVSGDKIYYFDETGAVKDTSKVDADKSSSAVSQN 209
Qy 232 ---PSINRAVDSSEASFELIDGYLTADSWTRPASIIKQGVYVWQASTADFPRLMAWVP 288
Db 210 ATIFAANNRAVTSAKNFEAVDNYLTADSWTRPKSILKDGKTWTESGDKDFFRLMAWVP 269
Qy 289 NVDTQVNYLYNYSKVFNLDAKYSSSTDQKFTLVKAAKDIQIKIEQIKQAEKSTQWLERETS 348
Db 270 DTETKRNVTYNNKVGVIDKTYTAETSQADLTAAAEALVQAREIQETKITSENNNTKWLREALS 329
Qy 349 AFVKTPQOWNKETEN-YSGKGGEHLQGGALLYVNDV-RTPWANSDYRLRLNRTATNQGT 406
Db 330 AFVKTPQOWNGESEKPY-----DDHLQNGALLFDNQDTLTPDTQSNRYLLNRTPTNQGS 384
Qy 407 IDKSILDSQSDPNHMGGFDFLLANDVLSNPVQAEQLNQIHYLMNWSIYVNGDKDANPD 466
Db 385 LDSRFTYNNPNDP--LGGYDFLLANDVDNSNPVQAEQLNWLHYLLNFGSIYANDADANPD 442
Qy 467 GIRUDAVDNDADMLQLYNTYPREYGYVKNSEANALAHISVLEAKSLNNDHNKDTDGAA 526
Db 443 SIRUDAVDNDADLLQISSDYLKAAAYGDKNKNANNHVSIVEAWSNDNPTPLDHDGDNL 502
Qy 527 LAMENKQRLALLFSLAKPIKERTPAVSPLYNNTFNTORDEKTDWINKDGSKAYNEDGTV 586
Db 503 MNMDNKFRLSMLWSLAKPLDKRS-GLNPLIHNSLVDRVDDR-----EVEIV 548
Qy 587 KQSTIGKNEYKYGDASGNYVFIIRAHNNVQDIIABIIKKEINPKSDGFTITDAEMKQAPE 646
Db 549 P-----SYSFARAHDSVQDIIIRDIKAEINPNPSFGYSFTQEEIEQAPK 592
Qy 647 IYNKMLSDSKKYTLNNTPAAYAVMLONMETITRYVYGDLYTDDGHYMETKSPYDITVYN 706
Db 593 IYNEDLKKTKKTYHYNVPLSYTLLLTNKGSIPIRVYVYGDYFDDGQYMWANKTVNYDAIES 652
Qy 707 LMKSRIKVYSGQAOQSYWLPDTPDGKMDNSDVELYRTNEVYTSVRYGKDOIIMTANDTEGSKY 766
Db 653 LLKARKKVVSGQOAMQNYQI-----NGEILTSVRYGKGAALKQSD-KGDAT 697
Qy 767 SRTSQVTLVANNPKLNDQSAKLVNEMGKIHANOKYRALIVGTADGINKFTSSDADAIAA 826
Db 698 TRTSGVGVVMGQPNFSLDGKV-VALNMGAAHANQBYRALMVSTKDG VATYATDADASKA 756
Qy 827 GYVKETDSNGVLTFGANDIKGYETFDMSGFVAVVYVPGASDNQDIRVAPSTAKKEGELT 886
Db 757 GLVKETDENGYLFLNDDLKGVANPQVSGFLQVWPVPGAADQDQDIRVAASDFASTDGK-S 815
Qy 887 LKATEAYDSOLLYEGFSNFPQTIPDGSDFPSVYNNRKAENVDLFSKSGVTSFEMAPQFVSA 946
Db 816 LHQDAAMDSRVWFEGFSNFPQSP--ATKBEETVNNVIANNVDKFSVNGIITDFEAPQVSS 873
Qy 947 DDGTFLDSVIQNGYAPADRYDLAMSNNKNGYKSKEDRLDALKALHAKAGIOIAIDWPDQYI 1006
Db 874 TDGQFLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAKGLVMADWVPDQMY 933
Qy 1007 QLPKGVEVVTATFTDAGAKRIADAIIDHSLYVANSKSGKDYQAKYGGSEFLABELKAKYPEM 1066
Db 934 TFPKQEVVTVTRTDKFGKPIAGSQNHSLYVYDVTXSSGDDYQAKYGGAFDELKELKYPEL 993
Qy 1067 PKVNMISTGKPIDDSVKLGKQWKAERYFNGCTNVLERGVYVLSDEATGKYFTVTKEGNFIP 1126
Db 994 FTKKQISTGOALDPSVKIKQWSAKYFNGSNILGRGADYVLSQVSKNYFNVASDITFLPS 1053
Qy 1127 QLTGKEKVTITGSSDGKITTY-FGTSGTQAKSAFVTFNGNTYFFDARGHMYTNSEYSPNG 1185
Db 1054 SLGLK-VWESGIRYDGKGYVYNSSATGDQVKASFITEAGNLNYFFGKDGVMYVGAQ-TING 1111
```


QY 1186 KDYRFLPGLMISNAFYIDANGNTYLYNSGQMYKGYTKYFDVSETDKDGKSKSVKVKFR 1245
Db 1112 AN-YFLENGTALRNTIYTDAQNSHYIANDGKRYENGYOQF-----GND-----WR 1157
QY 1246 YFTNEGVMAGVTVIDGFTQYFGBDGFQAKDK-LVTFPGKTYFYDAHTGNGIKDTW-RNI 1303
Db 1158 YF-KDGNMAGLTVTVGNVQYFKDGVQAKDKIIVTRDGRYFDDQNGNAATNTFIADK 1216
QY 1304 NGKYYYFDANGVAATGAQVINGQKLYFNEDEGSQVKGGVVKNVADGTYSKYKEGFGELVTNE 1363
Db 1217 TGHWYLLGKDGVAVTGAQTVKQKLYFEANGQQVKGDFVTSDEGKLYFYDVSDDMWTD 1276
QY 1364 FPTTDGNNVYAGANGKTVTGAQVINGQHLFYFNADGSQVKGGVVKNVADGTYSKYKNASTGE 1423
Db 1277 FIEDKAGNWFYLGKDGAAVTAQVIRGQKLYFKANGQQVKGDIIVKGTGDKIRIYDAKSGE 1336
QY 1424 RLTFNE-----PFTTG 1433
Db 1337 QVFNKTVKAADGTYVIGNDGVAVDPSVVKGTQFKDASGALRPNYLNKGQLVTGSGWYETA 1396
QY 1434 DNNWYIYGANGKSVTGEVKIGDDTYFFAKDGKQVKGQTVSAGNGRISYYYGDSGKRAVST 1493
Db 1397 NHDWVI-QSGKALTGEQTINGQHLFYKEDGHQVKGQVLTGTDGKVRYYDANSDDQAFNK 1455
QY 1494 WBIQFGVYVYFDKNGLA 1511
Db 1456 SVTVNGKTY-YFGNDGTA 1472

RESULT 9

US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; PRIOR FILING DATE: 2004-03-09
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match 41.9%; Score 3325.5; DB 5; Length 1590;
Best Local Similarity 45.0%; Pred. No. 7.2e-172;
Matches 710; Conservative 239; Mismatches 456; Indels 173; Gaps 33;
QY 1 MENKHYLKHKKQWHTTAVASVA-LATVGLGLSVTTSSVSADETQDKTVTQSGTGA 59
Db 1 MEKNVRFKMHVKRVRTVLSVASATWLASALG-----ASVASAD-----TDTASDD----- 46
QY 60 SLVTSPEAKTEADKRTNTEADVLTPAKETNAVETATTTTQATAEAAATTATTADVAVAA 119
Db 47 -----SNQAVVTGDTNNQATD-----QTSIAATAT-----SEQASATDAATQASAA 90
QY 120 VPNEAVVTTDAPAVTTEKAEQPAVKAEVVN-----TEVKAPEAAALKDSEVEAAL 171

Db 91 EOTQGTASTDTAAQTITNANE-AKVVPTEENENOGFTDEMLAEAKNVATAESDISPSDLA 149
QY 172 SLKNIKNIDGKIYVNVNEDGSHKENFALTNGQLLYFGKDGALTSSTSYSTPTPTTNIVDG 231
Db 150 KMSNVQVQDKGIYVDDGNNVKNFAVSVDKIYIFDETGYKDTQKVDADKSSVASQN 209
QY 232 ---FSINNRAYDSSEASFELIDGILTADSWYRPASIIKDGVTWQASTAEDFRPLLMAWP 288
Db 210 ATIFAANNRAYSTSAKNFEAVDNYLTADSWYRPKSIKDKGKTWESGKDDFRPLLMAWP 269
QY 289 NVDTVNYLYNYSKVFNLDKAYSSTOKETLKVAAKDIQIKIBQIKQAEKSTOWLBRETTIS 348
Db 270 DTETKENYNNKVVIGIDKTYTAETSOADLTAAAEVLQARIBOKITSENNTKWLREASIS 329
QY 349 AFVKTOPQWNETEN-YSKGGGEDHLQGGALLVNDV-RTPWANSYRRLNRTATNOTGT 406
Db 330 AFVKTOPQWNGSEKPY-----DDHLQNGALLFDNOTDLPDQTSYRLLNRTPTTQGTG 384
QY 407 IDKSIILDEQSDPNHMGDFDILLANDVDLSNPVQAEQLNQIHYLMNMGSTVMGDKDANFD 466
Db 385 LDSRFTYNDP--LGGYDFLLANDVNSNPVQAEQLNWLHLLNFGSIYANDADANFD 442
QY 467 GIRVDAVDNVDMLQLYTNVREYGVNKSANALAHISVLEAWSLNDNHNNDKTDGAA 526
Db 443 SIRVDAVDNVDADLLQISSDYLKAAYGIDKNNKNANNHVSIVEAWSNDNTPYLHDDGDNL 502
QY 527 LAWENKQRIALLFSLAKPIKERTPAVSPLYNNNTFNTQDEKTDMDINKGSKAYNDDGTV 586
Db 503 MNNDNFRSLMLWSLAKPLDKRS-GLNPLHNSLVDRVDDR-----EVETV 548
QY 587 KQSTIGKYNEKYGDASGNTYFIRAHNDNNVDIIAEIIKKEINPKSDGFTTDAEMKQAFE 646
Db 549 P-----SYSFARAHDSQVODIIRDIKAEINPNSFGYSFTQSEIQAFK 592
QY 647 IYNKMLSSDKKTYLNNIPAAVAMLQNMETTRVYVGLYTDGHHMETKSPYYTIVN 706
Db 593 IYNEDLKTKDKYTHYVPLSYTLTLTKNGSIPRVYVGMFTDDGGYMAKNTVNYDAIES 652
QY 707 LMKSRIKYVSGGQORSYWLPTDGRKNDSDVLYRNEVYTVSYRYGKDIWANDTSGSKY 766
Db 653 LUKARKYVSGGQAMQNYQI-----GNGEILTYSVRYGKALKQSD-KGDAT 697
QY 767 SRTSGQVTLVANNPKLNLDSAKLVNEMGKIHANQYRALIVGTADGIKNFTSDADAIAA 826
Db 698 THTSGVGVVGMGNQPNFSLDGKV-VALNMGAAHANQYRALVSTKDGVTATYATDADASKA 756
QY 827 GYVKETDSNGVLTFGANDIKGYETFDMSGFVAWVVPVGSADNQDIRVAPSTEAKGEGELT 886
Db 757 GLVKRTDENGILYFLNDDLKGVANQVSGFLQVWVPVGAADDDQDIRVAASDTASTDGK-S 815
QY 887 LKATEAYDSOLYIEGFSNFOITPDGSDPSVYTNRKIAENVDLFKSGVTSFEMAPQFVSA 946
Db 816 LHQDAAMDSRVNMFEGFSNFSQSF--ATKBEETVNVVIANNVDFVSGWITDFEMAPQFVSS 873
QY 947 DDGTFELDSVIQNGYAFADRYDLAMSKNNKYGSKEDRLDALKALHAGIQAIDAWDPDQIY 1006
Db 874 TDGQFLDSVIQNGYAFTRDYDLGMSKANKYGTADQLVKAIKALHAGLKVMAWDWPDQMY 933
QY 1007 QLPGKEVVTATTDGAGRKIADAIIDHSLYVANSKSKGDKYQAKYGGEFELAEKAKYPEM 1066
Db 934 TFPKQEVVTVTRTDKFGKPIAGSQINHSLYVDTTKSSGDDYQAKYGGAFDELKEKYPEL 993
QY 1067 PKVNMISTGKPIDDSVKLKQWAEYFNGTNVLERGVGVYLSDBEATGKYFTVTKGNFIFPL 1126
Db 994 FTKQISTGQAIDPSVKIKQMSAKYFNGSNIILGRGADYVLSDBQVSNKYNFVASDITLFLPS 1053
QY 1127 QLTGKEKIVITGSSSGKGIITY-FGTSGTQAKSAFVTENGNTYTFDARGHMVNSEVSPNG 1185
Db 1054 SLLGK-VVESGIRYDQKGYIYNSSATQOVKASFTIEAGNLVYFGKDGWTVTAQ-TING 1111
QY 1186 KDYRFLPGLMISNAFYIDANGNTYLYNSGQMYKGYTKYFDVSETDKDGKSKSVKVKFR 1245

```
Db 1112 AN-YPFLENGTALRNTIYTDAQNSHYYANDCKRYENGYYQF-----GND-----WR 1157
Qy 1246 YTNBEGWMAKVTVIDGFTQYFGEDEGFOAKOK-LVTFKGTYYFPAHTGNGIKDTW-RNI 1303
Db 1158 YF-KDGNMAVGLTVDGNYQYFDDKGQVAKDIIVTRDGKRVYFDQHNGNAATNTFIADK 1216
Qy 1304 NGKWYYFDGVAAGTAQVINGQKLYFNEDEGSOVKGVVKNADGTVSKYKEGFGELVTNE 1363
Db 1217 TGHYYLKGQGVAVTGAQTVGQKLYFEANGQGVKGFVTSDEGKLYFYDVDSGDMWDT 1276
Qy 1364 FETTDGNWYYAGANGKTVTGQVINGQKLYFNADEGSOVKGVVKNADGTVSKYNASTGE 1423
Db 1277 FIEDRAGNWFYLGKDGAAVTGAQTVGQKLYFKANGQGVKGDIVKGTGDKIRYYDAKSGE 1336
Qy 1424 RUTNE-----FPTTG 1433
Db 1337 QVFNKTVKAADGKTVYVIGNDGVAVDPVVKGOTFKDASGALRFYNLKGQLVTSQWYETA 1396
Qy 1434 DNNWYVIGANGSVTGEVKIGDDTYFFAKDGQVKGOTVSAGNGRISYYVYDGSQKRAVST 1493
Db 1397 NHDWVYI-OSGKALVTGEQINGOHLYFKEDGHQVKGQLVGTGDKVRYDYDANSQQAQFNK 1455
Qy 1494 WIEIQPVVYVYFDXGLA 1511
Db 1456 SVTVNGKTY-YFGNDGTA 1472

RESULT 10
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 40.3%; Score 3196.5; DB 3; Length 1475;
Best Local Similarity 45.2%; Pred. No. 6.7e-165;
Matches 695; Conservative 233; Mismatches 472; Indels 139; Gaps 31;

Qy 1 MENKIHKLKVKQKWTVIYAS--VALATVLGGLSVTTSSVSADFTQDKTVTQNSGTT 58
Db 1 MDKVRVYKLKRVKRVTVSVASAVMTLTLSGL-----VKADSNESKQISNDNNTS 54
Qy 59 ASLVTSPKATKADKKTNTKEADVLTPAKETNAVETATTNTTQATBAATATTADVAVA 118
Db 55 -----VVTANBESNVITEATSKQEAASSQTNHVTVTSSTSS 91
Qy 119 AVPNKEAV---VTTDPAVTTKEABEQPATV---KAEVVNTEVKAPEA---ALKDS--- 165
```



```
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-35

Query Match          38.7%; Score 3070; DB 4; Length 1375;
Best Local Similarity 46.2%; Pred. No. 4.6e-158;
Matches 661; Conservative 223; Mismatches 412; Indels 134; Gaps 33;

QY 1 MENKIHKLVKVKQWVITAVASVALATVGLSVTTSSVSADETODK--TVTQNSGTT 58
DB 1 MEKKVRFKLRKVKRWVTSIAS-AVVTI---TSLGSLVKADSTDRQQAATESQ---- 52

QY 59 ASLVTSPEATKEADKGTNTKEADVLPAKETNAVEATTNTQATAEAATATTADVAVA 118
DB 53 ASLVTTSEAAKETLTATDTSTATSATSQPTATVTDNVSTTN-QST---NTTANTANFVVK 108

QY 119 AVPNKEAVTTDA-PAVITTEKAEQ-----PATVKAENVNTEVKAPE-- 159
DB 109 PTTTSEQAQTDNSDKIITTSKAVNRLTATCKFVPANNNTAHPKTVTDKIVPKIKGKLUK 168

QY 160 --AALKDSEVEAALSUKNTKIDGKYVYNNEDGSHKENFAITVNGQLLVFGDGALTSSS 217
DB 169 QPSSLSQDDIALGONVNRKNGKYYKYKEDGTLOKNVALMINGKTFPDETGALSNTT 228

QY 218 TVSFTEPTGTT--IVDGFSSINRAYDSSEASFELIDGYLTADSWYRPASIKDGVWQAST 275
DB 229 LPSKKGNIITNDNTNSFAQVNVSTDVANFEHVDHYLTAESWYRPKYILKDKCTWQST 288

QY 276 AEDFRLLMAWPNDVTQVNLNYSKVFNLDAKYSSTDKQETLVKAAKDIQIKTEBQTK 335
DB 289 EKDFRELLMTWPDQETQRYVYNNAAQLGIHQTYNTATSPQLNLAAQTIOIKTEBKIT 348

QY 336 AEKSTOWLRETTISAFVKTOPWNKENTENYSKGGEDHLOGGALLVYVNDSR--TPWANSDYR 394
DB 349 AEKNTWNLQRTISAFVKTSQAMNSDSEK-----PFDDHLQKGALLYSNNKSLTSQANSYR 404

QY 395 RLNRATNTGTITDKSILDEQSDPNH-----MGDFDFLLANDVDVLSNPVVAEQLNQIHY 449
DB 405 ILNRTPTNQTG-----KKDPRTADRTIGGYEFLLANDVDNSNPVVAEQLNWLHF 455

QY 450 LNNWGSIVMGDKDANFDGIRVDAVNDVADMLQLYTNFYREYVYGVNKSANALAHISVLE 509
DB 456 LMNFGNIYANDPDANFDSIRVDAVNDVADLLQIAGDYLKAAGTIHKNDKAANDHLISLE 515

QY 510 AANSLNDHYNDKTDGAALAMENKORLALLFSLAKPIKERTPAVSPLYNNTP--NTTQDRBK 568
DB 516 AMSYNDTPYLHDDGDMNIMNDRLSLUSLYSLAKPLNQRS--GMNPLITNSLVNRTDDNAE 574

QY 569 TDWINKDGSKAYNEDGTVAQSTIGKYNKYGDASGNYVFIKRAHDNNVQDIIAEIIKKEIN 628
DB 575 T-----AAPSYSFIRAHDSVQDLIRNIIRTEIN 604

QY 629 PKSDGFTIIDAEMKQAFEYNKDMLSDDKYITLNNIPAAAYAVMLQNMETITRVIYGDLYT 688
DB 605 PNWVGSYFTTEIKKAFEYNKDLLATEKKYTHYNTALSYALLLTNKSVPVRYGDMFT 664

QY 689 DGHVYMETKSPYVDTIVNLMKSRISKVSGQAORSYWLPTDGMNDNSDVELYRTNEVYTS 748
DB 665 DGQYWAHTIYNEALETLKARIKIVSGQAARNQ-----QVGN-----EITS 710

QY 749 VRYGKDIMTANDEGSKYRSTSQVTLVANNPKNLDSQAKLVNEMGKIHANQKYRALIV 808
DB 711 VRYGKALKATDTI--GDRTTTRTSGVAVIEGNNPSRLKASDRVVVNNNGAAHKKQAYRPLL 769

QY 809 GTADGINKTSDADATAAGVYKTEDSNGVLTFGANDIKGYETDFMSGFFVAVVVPVCASDN 868
```

```
Db 770 TTDNGIKAYHSQDE--AAGLVRYTNDRGELIFTADIKGYANPQVSGYLGWVPVGAAD 827
QY 869 QDIRVAPSTEAKKEGELTKATEAYDSQLIYEGFSNFQIPTDGSDFSVVTVNRKIAENVDL 928
Db 828 QDVRVAASTAPSTDGK-SVHQNAALDSRVMPGFSNFQAF--ATKKEEYTNVVIKKNVDK 884
QY 929 FKSXGVTSPFEMAPQPVASADDGTFLDSVTQNGVAFAPADRVDLAMSNNKNGYSGKESDLRLAKA 988
Db 885 FAEWGVTDFEMAPQVSVSSDGSFSLDSVIQNGVAFTRYDGLGISKPNKYGTADDLVKAIFA 944
QY 989 LHKAGIOAIADWVPDQIYQLPGKEVVVTRTDTGACRKIADAJIDHSLYVANSKSSKQVQ 1048
Db 945 LHSKGIKVMADWVPDQMTALPEKEVVTATRVDKYGTVPAGSQIKNTLYYVVGKSSGQDQ 1004
QY 1049 AKYGEFELAEAKAKYPEMFKVNMISTGKPIDDSVKLKQWKAEYFNGTNNLGRGVGLSD 1108
Db 1005 AKYGGAFLEELQAKYPELPARKQISTGVPMDFSVKIKQWSAKYFNGTNNLGRGAGYVLKD 1064
QY 1109 EATGKYFTVTKEGNFIPIQLT-----GKEKVITGFSDDGKGIYFTGTSGTQAKSAFVTFNG 1164
Db 1065 QATNTYFSLVSDNTFLPKSLVNPNHGTSSTSVTGLVFDGKGYVYVYSTSGNQAKNAFISLGN 1124
QY 1165 NTYYFDARGHMVTNSEYSPNGKDVYRFLPNGLMISNAFYIDANGNTYL--YNSKGOMYKG 1222
Db 1125 NWYFDNNGYMTVGAQ-SINGANYI-FLSNGIQLRNALY--DNGNKVLSYNGDGRRYEN 1180
QY 1223 GYTFEDVSETDKDGKESKVKVPFRYFTNEGVMKAGVTVIDGFTQYFGEEDGFOAKDLV-TF 1281
Db 1181 GYTLF-----GQQ-----WRYFQN-GIMAVGLTRVHGAQVQYDPASGFOAKGQFITTA 1226
QY 1282 KGTYYFDDAHTONGIKOTW--RNINGKYYTFDANGVAATAQVINGOKLYFNEDGSGVQKG 1340
Db 1227 DGKLYRFDKSGNQISNRFVRNSKGEWFLFDHNGVAVTGTVTFTFNGQRLYFKFNGVQAKGE 1286
QY 1341 VYKNADGTVSKYKEGFGELVTNEFFTTDGNVWVYAGANGKTVTGQVING 1390
Db 1287 FIRDANGYLRYYPDPSNGNEVRNFRVNRNSKGEWFLFDHNGIAGVITGAKVVNG 1336
```

RESULT 15

```
US-10-797-821-35
; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10797,821
; PRIOR FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-35
```

Query Match 38.7%; Score 3070; DB 5; Length 1375;
Best Local Similarity 46.2%; Pred. No. 4.6e-158;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 8.6381 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-40
Perfect score: 7928
Sequence: 1 MENKIYKLUHKVKKQWTTIA.....PGVYVYFDKNGLAYPRVLN 1518

Scoring table: BLOSUM62

Gapbp 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_New.*
1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pap.*
2: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pap.*
3: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pap.*
4: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pap.*
5: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pap.*
6: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pap.*
7: /cgn2_6/ptodata/1/pubppaa/US11_NEW_PUB.pap.*
8: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	522	6.6	2710	7 US-11-051-453-41	Sequence 41, Appl
2	367	4.6	2367	7 US-11-051-453-42	Sequence 42, Appl
3	338.5	4.3	1992	7 US-11-013-759-3	Sequence 3, Appli
4	338.5	4.3	1992	7 US-11-013-759-13	Sequence 13, Appl
5	338.5	4.3	2047	7 US-11-013-759-4	Sequence 4, Appli
6	338.5	4.3	2047	7 US-11-013-759-7	Sequence 7, Appli
7	322	4.1	396	7 US-11-022-562-228	Sequence 228, App
8	316.5	4.0	2053	7 US-11-013-759-9	Sequence 9, Appli
9	301	3.8	726	7 US-11-052-554A-217	Sequence 217, App
10	299	3.8	2314	7 US-11-013-759-11	Sequence 11, Appl
11	280.5	3.5	619	7 US-11-052-554A-229	Sequence 229, App
12	276.5	3.5	701	7 US-11-052-554A-231	Sequence 231, App
13	267	3.4	744	6 US-10-873-528-184	Sequence 184, App
14	255.5	3.2	1647	7 US-11-052-554A-260	Sequence 260, App
15	251	3.2	693	6 US-10-873-528-185	Sequence 185, App
16	246.5	3.1	1767	7 US-11-052-554A-372	Sequence 372, App
17	244.5	3.1	5024	6 US-10-793-626-2964	Sequence 2964, Ap
18	236	3.0	1155	6 US-10-793-626-1780	Sequence 1780, Ap
19	232.5	2.9	5291	7 US-11-052-554A-281	Sequence 281, App
20	226	2.9	627	6 US-10-873-528-191	Sequence 191, App
21	225	2.8	2902	7 US-11-052-554A-91	Sequence 91, Appl
22	224	2.8	332	6 US-10-873-528-190	Sequence 190, App
23	224	2.8	1588	7 US-11-052-554A-280	Sequence 280, App
24	221.5	2.8	1263	6 US-10-485-517-127	Sequence 127, App
25	220	2.8	690	7 US-11-052-554A-232	Sequence 232, App

26	219	2.8	338	7 US-11-052-554A-228	Sequence 228, App
27	211	2.7	1571	7 US-11-052-554A-2	Sequence 2, Appli
28	207.5	2.6	1296	6 US-10-615-668-3	Sequence 3, Appli
29	207	2.6	1448	6 US-10-485-517-212	Sequence 212, App
30	207	2.6	2233	6 US-10-873-528-2	Sequence 2, Appli
31	206.5	2.6	1694	7 US-11-052-554A-83	Sequence 83, Appl
32	204	2.6	3194	7 US-11-052-554A-90	Sequence 90, Appl
33	202	2.5	1579	7 US-11-052-554A-9	Sequence 9, Appli
34	201.5	2.5	658	6 US-10-873-528-17	Sequence 17, Appl
35	201.5	2.5	677	6 US-10-873-528-155	Sequence 155, App
36	198.5	2.5	2340	7 US-11-052-554A-171	Sequence 171, App
37	196	2.5	1417	7 US-11-052-554A-8	Sequence 8, Appli
38	196	2.5	2551	7 US-11-052-554A-368	Sequence 368, App
39	195.5	2.5	2399	7 US-11-052-554A-92	Sequence 92, Appl
40	193.5	2.4	340	6 US-10-873-528-188	Sequence 188, App
41	193.5	2.4	1706	7 US-11-052-554A-176	Sequence 176, App
42	193	2.4	318	6 US-10-873-528-192	Sequence 192, App
43	193	2.4	318	7 US-11-052-554A-233	Sequence 233, App
44	192	2.4	1751	7 US-11-103-957-45	Sequence 45, Appl
45	191.5	2.4	1889	7 US-11-102-476-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-11-051-453-41
; Sequence 41, Application US/11051453
; Publication No. US20050287150A1
; GENERAL INFORMATION:
; APPLICANT: AMBROSINO, DONNA
; APPLICANT: BABCOCK, GREGORY J.
; APPLICANT: BROERING, THERESA
; APPLICANT: GRAZIANO, ROBERT
; APPLICANT: HERNANDEZ, HECTOR JAVIER
; APPLICANT: LOWY, ISRAEL
; APPLICANT: MANDELL, ROBERT
; APPLICANT: MOLRINE, DEBORAH
; APPLICANT: THOMAS, JR., WILLIAM D.
; APPLICANT: ZHANG, HUI-PEN
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
; FILE REFERENCE: MJI-001
; CURRENT APPLICATION NUMBER: US/11/051,453
; PRIOR FILING DATE: 2005-02-04
; PRIOR APPLICATION NUMBER: 60/542,357
; PRIOR FILING DATE: 2004-02-06
; PRIOR APPLICATION NUMBER: 60/613,854
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: Patentin Ver. 3.3
; SEQ ID NO 41
; LENGTH: 2710
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 6.6%; Score 522; DB 7; Length 2710;
Best Local Similarity 20.1%; Pred. No. 2.2e-20;
Matches 318; Conservative 194; Mismatches 491; Indels 576; Gaps 71;
QY 176 IKNDGKYVNVNEDGSHKENFA--ITVNGQLYFGKDGALTSSTVSFTPTGTTNIVDGRS 233
1438 INSLNTEKINTGLDLSKNIAVNTYDSENNKYFGAISTKSQSIHYKKDSKNILEFYN 1497
QY 234 INNRAYDSEASFELIDGYLTADSWYRPASIIKDGVTWOASTAEDFRPLLMAMPNVDTQ 293
1498 DSTLEFNSKDFTAEDINVFMKD-----1520
QY 294 VNYLNMVKVFNLDKYSTDKQETLKVNAKDIOIKIEOKIAEKSTQWLRETISA----349
1521 ---INTITGKYVD---NNTDK-----SIDFSISLVSKNQVKNGLYNESVSYLD 1567

QY 417 DPNHMGDFELLANDVDLSNPVQAEQLN-----QIHYLMNWSIVMGDKDANFDGIRVDA 472
Db 869 D-----ALVNAKDI-----AENLNTLAKEIH-----TTKGTADTALQTFTVKK 906
QY 473 VD-NVDADMQLQYTYFREYYGNKSEANALAHISVLEAWSLNDHNDKTDGAALAMEN 531
Db 907 VDENNADANAIT-----VQKVNANOVNTLTAK-----GENGLNIKTD----- 946
QY 532 KQRLALFLSLAKPIKERTPAVSPLYNNFTNTQDEKTDWINDKSKAYNEDGTVKQSTI 591
Db 947 -KNGTVTFGI-----NTTSLRAGKST--LNDGELSINKPTCS-EQIQV 986
QY 592 GKNEYKGDASGNVFIIRAHNNVDIIAEIIKKEINPKSDGFTITDAEMKQAFIYNKD 651
Db 987 GADGVKPAKVNNGVWAGIDGTTT-----ITRDEI-----GFTGTNGSLDKSKPHLSKD 1036
QY 652 MLSSDKKYTLNIPAAVAVMLQNMETITRVYVYDLYTDGHHMETKSPYYD---TIVNLM 708
Db 1037 GINAGK-----KITNQSDEI-AQNSHDVATGKIIDLTKELENKI 1077
QY 709 KSIKIVSGQQAQSRWMLPTDGRMNSDVELYRTNEVYTSVRYGKDIIMTANDTEGSKYSR 768
Db 1078 SSTAK-----TAQNS-----LHEFSVADEQGNFTVSNPYSSYDTSK 1114
QY 769 TSGQVTLVANN-----PKNLNDQSACL-----NVEMGKIHANQYKRALIVG 809
Db 1115 TSDVITPAGENGITTKVNGWVRVIGIDQTKGLTTPKLTGVNNGKGIIVDSQNGQNTITG 1174
QY 810 TAGIKNFTSDADAI-----NAGVYKETSND-----GVLTFGAN-----DIKGYET 850
Db 1175 LSTLANVTNDKGSVRTTEQNTIIOBKEDKTRAASIVDVLVSAGFNQNGEAVDFVSTYDT 1234
QY 851 FDMSGFVAVVVPVGCASDQ-----DIRVAPSTEAKGEGEL-----TLKATEAYDSOLI 898
Db 1235 VNFADGNATTAKYTDTSKTSKVVDVNVDDTIIIEVKDKLGVKVTTLTSTGTGANKFA 1294
QY 899 YEGFSNFQIPDGSDFSVYTRKIAENVDLFSKWSGVTSEMAPQFVSADDTGLDLSVIQN 958
Db 1295 LSNQATGDALVKASDIVAHLN---TLSGDIIQTAKGASQANNSAGYVDADGNKVI----- 1345
QY 959 GYAPADRYDLAMSKNKY--GSKEDLRDALKALHKAGIQAIADVPDQIYQLPGKEVWTA 1016
Db 1346 -----YD-----STDNKYIQAQKNDGTVDKTKEVAKDKLVAQAQTPDGTTLAQNMVKSIVNK 1395
QY 1017 TRTDGAGRIADAI IDHSLVANSKSGKDYQAKYGEFLAEK--AKYPEMPKVNMI 1074
Db 1396 EQVNDANK--QGINEDNAFVKGLEKASDNKTKNAAVTVGDLNVAQAQPLTFAGDTGTT 1453
QY 1075 GRPIDDSVKLKQWKAIFYNGTNVL--ERGVGYVLSDEATGKYFTVTKEGNFPIPLQLTKREK 1133
Db 1454 AKKLGETLTIKGGQTD---TNKLTNNIGVAVGTG---PTVK-----LAKDLTLNLS 1500
QY 1134 VITGFSS-DKGITYFGTSG--TOAKSAFTVFNNGTYTFDARGHMVTNSEYSPNGKDVYRP 1191
Db 1501 VNAGGFKIDDKGVSFVDSGQAKANTPVLISANG---LDLGGKVIISNVGKGTGKTDAAV 1556
QY 1192 ----LPGNIMLSNAFVIDANGNTYLYNSKGOMYKGYTKFDVSEYDQKESKVVYKPY 1246
Db 1557 QQLNEVRNLGLGNAGNDADGN-----QVNIADIKDQPNSGS----- 1594
QY 1247 FTNEGVMAGVTVIDG----FTQYFGEDGFQ-AKDKLVTFKGTYYFYFDAHTNGIKMDTW- 1300
Db 1595 -SSNRVTIKAGTVLGGKGNNDTEKLATGGIQQGVDK-----DGNANGDLSNVVV 1642
QY 1301 -----RNINGKYYIFDANGVAATAQAQVINGOKLYFNEDGSGQ---VKG--GV 1341
Db 1643 KTKQGSKKALLATYNAAGQNTYLTNNPAEAIIDRINEQGRIFRPHVNDGNQEPVVOGRNGI 1702
QY 1342 VKNADGTSKYKEGFGELVINEFF-----TTDGNWYVYAGAKVTGTAQ-----V 1387
Db 1703 DSSASGKHS--VAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDNAQ--ATGDQSAIGTGNV 1760

QY 1388 INQHLYFNADSGOVKGVVKNADGTYG--KYNASTGERLITNEF-----FTTGDNNWYYI 1440
Db 1761 VAGKHSIGAIDPSTVK-----ADNSYSGNNOFTDATQTDVFGVGNNTVTTESNSVAL 1814
QY 1441 GANGKSVTEGVKIGDDTYFPAKDGQKVGQTVSAG--NGRISYYVYDGSGRKRAVS 1492
Db 1815 GSN-----SAISAGTHAGTOAKKSDGTAGTTTGTAGATGTVKGFAGTAVGAVS 1862
RESULT 5
US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sabaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-4
Query Match 4.3%; Score 338.5; DB 7; Length 2047;
Best Local Similarity 19.5%; Pred. No. 1.4e-10;
Matches 315; Conservative 228; Mismatches 633; Indels 437; Gaps 73;
QY 33 LSVTSSVSADETQDKVTQTSNSGTTASLVTSF-----EATKEAD-----KRTNT 77
Db 589 IGVKTELSNDSGTSK-PSVKGSGTNNSLVTAHLASLYLNEVNRADSLQSFVKEEDD 647
QY 78 KEADVLTPAKETAETATTTNTQATAEAAATTATTADVAVAAPNKE--AVVTTDAPAVT 135
Db 648 DDANAITVAKD-----TTKNAGAVSILKLGKNGLTVATKKDGTVT 688
QY 136 TEKASEQPAATVKAENVTEVKAPEALAKDSEEAALSKNIK--NIDGKYYVYNEGDSHK 193
Db 689 FGLSQDSGLTIQKSTLND-----GLTVKDTNEQIQVGANGIKFTNVNG-----SNPGTGI 739
QY 194 ENFAITVNGQLLYFGKDGALTSSSTY-----SFTPGTTNIVD--GFSINNRY----- 239
Db 740 ANTARITRDKIGFAGSDGAVDTNKPYLDDQDKLQGVNVTNTGINAGGKAITCLSPFLPS 799
QY 240 --DSSEASFELIDGYLTADSWRTPASIIKDGVTWQAETAEDPRLLMAWPNVDTOVNYL 297
Db 800 IADQSSRNIEL-----GNTIQDKDKSNAASIND-----ILNTGFLNKNNNPI 842
QY 298 NYMSKVFNLD--AKYSSTDKQETLVAAKDIQIKISQIOAKESKTQWLRETSFAVKTQ 356
Db 843 DPFVSTYDVIDFANGNATATVTHDTANK--TSKVYVDVNVDDTTHLT----- 888
QY 357 MNKETENYKSGGGEHLQGGALLYVNDSTRTPWANSDYRLNRNTATNTQGTIDKSLDEQS 416
Db 889 -----GTDNKKLG-----KTYKLNKTSANGTATATFNVNSSDE 923
QY 417 DPNHMGDFELLANDVDLSNPVQAEQLN-----QIHYLMNWSIVMGDKDANFDGIRVDA 472
Db 924 D-----ALVNAKDI-----AENLNTLAKEIH-----TTKGTADTALQTFTVKK 961
QY 473 VD-NVDADMQLQYTYFREYYGNKSEANALAHISVLEAWSLNDHNDKTDGAALAMEN 531
Db 962 VDENNADANAIT-----VQKVNANOVNTLTAK-----GENGLNIKTD----- 1001


```

1092 DB      GINAGGK-----KITNIQSGEI-AQNSHDAVTGCKIYDLKTELENKI 1132
709  QY      KSRIKVSQQAQSYWLPDGRWMSDVELYRTNEVTSVRGKDIMTANDTEGSKYSR 768
1133 DB      SSTA-----TAQNS-----LHEFSVADEGNNFTVSNVSSYDTSK 1169
769  QY      TSQVTLVANN-----PKNLDSQAKL-----NVEMGKIHANQYRALIVG 809
1170 DB      TSDVITFAGENGITTKVNGVVRVIGIDQTKGLTTPKLTGNNNGKGVIDSQNGQNTIG 1229
810  QY      TADGIKNFTSDADAI-----AAGVVKETDSN-----GVLTFFGAN-----DIKGYET 850
1230 DB      LSNLTANVTNDRKGSVRTTEQGNIIKDEKTRAASIVDVLVSAGFNLOGNGEAVDFVSTYDT 1289
851  QY      FOMSGFVAVVWVPGASDNO-----DIRVAPSTEAKKEGEL-----TLKATEAYDSQLI 898
1290 DB      VNPADGNATTAKYDTSKTSKVVDVNVDDTTIEVKOKKLGKVTITSTGTGAKFA 1349
899  QY      YEGFSNFQITPDGSDPSVYTNRKIAENVDLFKSWGVTSPFEMAPQFVSADDTGTFDLSVIQV 958
1350 DB      LSNQATGDALVRASDIVAHLN---TSLGDIQTAKGASQANNSAGYVDAGNKVI-----1400
959  QY      GYAFADRYDLAMSKNKY---GSKEDLRDALKALHKAGIOAIADWVPDQIYQLPKGEVVTVA 1016
1401 DB      -----YD-----STDNKYTAQKNDGTVDKTEVAKDKLVAQAQTPDGTTLAQMNVKSVINK 1450
1017 QY      TRTDGAGRIADAIIDHSLYVANSKSSGKDYQAKYGBEFLAELK--AKYPMPKVMNIST 1074
1451 DB      EQVNDANKK--OGINEDNAFVKGLEKAASDNKTNAAVTVGDNLNAVAQTPLTFAGDTGT 1508
1075 QY      GKPIDDSVKLQWKABEYFNGTNVL--ERGVGYVLSDEATGKYFTVTKEGNFIPQLTGRKEK 1133
1509 DB      AKKLGELTIKGGQTD---TNKLTNNIGVAGTDG---FTVK-----LAKDLTNLNS 1555
1134 QY      VTGPFSS--DGKITYPEGTS--TOAKSAFTVFNCTNYTFDARGHMTVNSYSPNGKDVYRP 1191
1556 DB      VNAGGTKIDDKGVSVFVDSGQAKANTFVLSANG---LDLGGKVISNVGKGTDKTDAANV 1611
1192 QY      -----LPGMILSNAPYIDANGNTLYNSKGMQYKGYTKFDVSETDKDGKESKVKVPY 1246
1612 DB      QQLNEVRNLGLGNAGDNADGN-----QVNIADIKDPNSGS-----1649
1247 QY      FTNEGVMAGKVTVIDG---FTQYFGEDEGFO-AKDKLVTFKGTYYFDHTGNGIKDTW- 1300
1650 DB      -----SSNRTVIRKAGTVLGGKNDTEKLATGIGIQGVK-----DGNANGDLSNVWV 1697
1301 QY      -----RNNGKYYFFDANGVAATCAQVINGOKLYFNEBDSQ---VRG--GV 1341
1698 DB      KTQKDGSKKALLATYNAAGQTNVLTNNPAEIDRINEQGIREFHVDNGNQEPVQGRNGI 1757
1342 QY      VKNADGTSYKKEGRELVTNEFF-----TTDGNVWYAGANGKTVTGAQ-----V 1387
1758 DB      DSSASGKHS--VAIGFOAKADGAAVAIGRQTOAGNQSTIAIGDNAQ--ATGDQSIATGTGNV 1815
1388 QY      INGOLHYFNADGSQVKGKGVVKNADGYS--KYNASTGERLTNEF-----FTTGDNNYYI 1440
1816 DB      VAKHSAGLGDPSYVK-----ADNSYSGVNNNOFTDQTDVFGVGNNTVTSNSVAL 1869
1441 QY      GANGSKSVTGEVKIGDDTYFFPAKDGKQVKQTSYAG--NGRISYYGDSGGRAYS 1492
1870 DB      GSN-----SAISAGTHAGTQAKKSDGTAGTITTTAGATGTVKGFAGTAVGAVS 1917
```

RESULT 7

```

US-11-022-562-228
; Sequence 228, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
```

RESULT 8

```

US-11-013-759-9
; Sequence 9, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MS:Jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
```

```

; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-11-022-562-228
```

```

Query Match      4.1%; Score 322; DB 7; Length 396;
Best Local Similarity 24.9%; Pred. No. 1.1e-10;
Matches 126; Conservative 53; Mismatches 164; Indels 164; Gaps 21;

QY      1033 HSLYVANSKSGKDYQAKYGG-----EFLAELKAKYPMFKV-----NMISTGKPIDDS 1081
DB      16 HKVMINLDSSEFYKWNSTEGSDPFLVRYLESNKKILQIRIKGILSNKTSFNKMSIDF 75

QY      1082 VKLKQWKABY-----FNGTNVLERG--VGYVLSDEATGKYFTVTKEGNFIPQLTGRKEK 1133
DB      76 KDIKLSLGYIMSNFKSFENSENELORDHLGFKIIDNKTYVYDEAS-----K 121

QY      1134 VTGFSGDKGITYTGTSGTQAKSAFTVFNCTNYTFDARGHMTVNSYSPNGKDVYRFLP 1193
DB      122 LVKGLININNSLFYDPDIESNLVTGWTINGKYYFDINTGAASTSYKILINGKHFY-FNN 180

QY      1194 NGIMLSNAPYIDANGNTLYNSKGMQYKGYTKFDVSETDKDGKESKVKVPYFTNEGVM 1253
DB      181 NGVMQLGVF-----KG---PDGFYFAPANTQNNIEGQAIIVY-----215

QY      1254 AKGVTVIDGFTQYFGEDEGFOAKDKLVTFKGTYYFD---AHTGNGIKDTWRNNGKMY 1309
DB      216 -----QSKFLTLNGKYYFDNDSKAVTG-----WQTDGKYY 248

QY      1310 PDAN-GVAATGAOVINGOKLYFNEBDSQVKGKGVVKNADG-----TVSKYNASTG 1368
DB      249 FNLTAEAAATGWTIDGKYYFN-----TNTSIAS- 278

QY      1369 GNVVYAGANGKTVTGAOVINGOHLYFNADGSQVKGKGVVKNADG-----TVSKYNASTG 1422
DB      279 -----TGYTIINGKHFYFNTDGI-MQLGVFKGPNGFYFAPANTDANNIEG 323

QY      1423 E--RLTNEFFTTGDNNWYIIGANGKSVTGEVKIGDDTYFFPAKDGKQVKQTSYAGN--- 1477
DB      324 QAIRQNRFLYLHD-NIYFNGNSKAVTGMQWQTINGNVVYFMPD-----TANAAAGGLPE 376

QY      1478 --RISYYGDSGGRAVSTWIEIQGVY 1502
DB      377 IDGVIYFFGVGVKA-----PGIY 395
```

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 9

; LENGTH: 2053

; TYPE: PRT

; ORGANISM: Moraxella catarrhalis

US-11-013-759-9

Query Match

Best Local Similarity 4.0%; Score 316.5; DB 7; Length 2053;

Matches 347; Conservative 225; Mismatches 696; Indels 445; Gaps 81;

Db 24 VALATVGLSVTTSSVSADETDKTVTQSNST---TASLVTSPEATKEADKRTNKE 79

Db 412 VKLAKELTGL---TSVSA--TNKIVTNTNNNAELQSGGLTFSPITGKTDK--TVVS 463

Qy 80 ADVLTPAKETNAVETATTNN--TQATAEAATATTADVAVAAPNKEAVV---TTDPAPVT 135

Db 464 IDGLKFTDNSIATGTTTRITKKIGPAGTNDGVDESKPYLDNEKLKVGNTSLNSSLT 523

Qy 136 TEKAEHQP-----ATVKAENVN-----TEVKAPAAALXDSEVEAALSILK 174

Db 524 VNNVTGNKQIQVANGIKFATVANNVANTSATVGTARITEEKIGFAGTNDGVDEQAPYLD 583

Qy 175 NIKNIDGKYVYVNEDESHKENFAIT--VNGQLLYFGKOGALTSSTSYSTPTGTTNIVDGF 232

Db 584 KERLKVGRVEIITDGSINAGNHKITGLTNG-----TANTDAVTIKQLDKAKP--TLNAGDGI 638

Qy 233 SINRAYDSSEASFELIDGLTADSW---YRPASIIKOGVT---WQASTAEDFRLPLMA 285

Db 639 SINSNNGDLVDSS-----GNITPTVNI SVKTTKLNSTNGTSGNNKFSVSNADHNNSLV-- 691

Qy 286 WNPVVDVTQVNYLNYMSKVFNLDKYSSTDKQETLKVAADKDIQK-----IEQKIQAEKSTQ 341

Db 692 ----TAKDLADYLNKV-----NETADSALPSPFKVQNGDNNNAITVVGKOT- 732

Qy 342 WLRETIASAFVKTPQPNKETEYSGGGEDHLQGGALLVYNDSTRTPWANSDYRLNRTAT 401

Db 733 -----NGKTFNLTCLKGEN-----GWNITTN 753

Qy 402 NOTGT-----IDKS-----ILBQSDPNHMGDFDILL----- 428

Db 754 RATGVTTFGIDQSNGLTTPKLTVGSDTNGNRLVIEQVPSADGNSKTKNIIKGLSPTLPSTA 813

Qy 429 ---ANDVDLSNPVQABQLN--QIHYLMWG--SIWVGDKDANFDGIRDAVDNADMLQ 482

Db 814 SPGRNIALGNTIEBKDKSNAASIDDLVNLNAGFNKNGKDKDFVS--TYDTVDFIDGNATT 872

Qy 483 LVTNY-----PREYGVNKSIA-----NALAHISV---LEAWSLNDNHNDKTDCA 525

Db 873 ATVTYDEANQTSKVAYDVNVDEKTIELTDGNGKKQLGVKTIKLTETSTNGNATTTSTDD 932

Qy 526 ALAMENKORALLFLSLAKPI-KERTPAVGPLYNNFTNTTQDEKTD-----WINKDGSKA 579

Db 933 HALVKASDIAGNLTALAEIHTTKGTANTALQTFVKKYVDENDKADDTNATITVGKGTSG 992

Qy 580 -----YNEDGTVKQSTIGKNEYKGDAGNVFIRAHNNVQDIIAEII 623

Db 993 KVNTLKLKGNGLIDIKTDKGTV---TFG-INTQSLKAGDSTTL---NNNGLSIKNTAS 1045

Qy 624 KKEINPKSGFTTDAEMQAPEIYNKMDLSSDKKYTLNNIPAAVAVMLQNNMETITRVY 683

Db 1046 NQOIQVGADGV-----KFAWNVNGVVGAG-----IDGTRITRDEI 1081

Qy 684 GDLTDDGHMETKSPYYDTI-VNLMSKRIKYSVGGQ-AQRSYWLPFDGMDNSDVELYR 741

Db 1082 G--FTGTNGSLDKSPHLSKDGINAGGKKITNIQSGEIAKNSHDAVTGGKIYDLKTEL-- 1137

Qy 742 THEVYTSVRYGKDIM---TANDEGSKY-----SRSSGQVTLVANN----- 779

Db 1138 ENKISSTAKTQNSLHFEVSVADEQGNFTVSNPYSSYDTSKTSVDITFAGENGITTKVKN 1197

Qy 780 --PKMLDQSAKL-----NVEMGKIHAKQYRALLIVGTADGINKFTSDADAI---- 824

RESULT 9

US-11-052-554A-217

; Sequence 217, Application US/11052554A

; Publication No. US20050288866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 217

Db 1198 GWRVGIDQTKGLTTPKLTVGNNGKGIVINSQNGQNTITGLSNTLANVTNDKGSVRTTE 1257

Qy 825 AAGYVKETDSN-----GVLTFGAN-----DIKGYETFDMSGGFVAVWVPVGCASDQ 869

Db 1258 QGNIKDEBKTRAAISVDVLSAGFNLCQNGEAVDFVSTYDVTNFGANGNTTAKAVTYDDTS 1317

Qy 870 -----DIRVAPSTEAKKEGEL-----TLKATEAYDSOLIVYBGFNSFQTI PDGSDPSVY 917

Db 1318 KTSKVVDVNVDDTTTIEVKDKKLGKVTTLTSTGTGANKFALSNQATGDALVKASDIAVAH 1377

Qy 918 TNRKIAENVDLFKSWGVTSEFMAPOFVSADDTGTFDLSVIONGYAFADRYDLAMSKNNKY- 976

Db 1378 LN-----TLSDGIQTAKGASQANNSAGYVDADGNKVI-----YD---STDNKKY 1418

Qy 977 -GSKEDLDALKALHAKHAGIQAIDWVPDQIYQLPQKEVVTATRTDGAQRKIADAIIDHSL 1035

Db 1419 QAKNDGTVDTKTEVAKDKLVAQAQTPDGLTACQMNKVSINKEQVNDANKK--QGINEDNA 1476

Qy 1036 VYANSKSGDKYQAKYGGEGFLAELK--AKYPEMPKVMNISTGKPIDDDSVKLGKQWKAIFYN 1093

Db 1477 FVKGLEKAASDNKTKNAAVTVGDLNVAQAQTPLTTFAGDTGTAKKLGELTITKGQTD--- 1533

Qy 1094 GTNVL-ERGVGVVLSDEATGYFTVTKEGNFIPQLTGKEKVIYTFSS-DGKGIYFGTS 1151

Db 1534 -TNKLTDDNNIGVAGTDG---FTVK-----LAKDLTNLNSVNAGGTKIDEGGISFVDAN 1583

Qy 1152 GTQAKSAFVTNGNTYYFDARGHVMYTNSEYSPNGKDVYRFLPNGIMLSNAFYDANGNTY 1211

Db 1584 G-QAKA-----NTPVLISA-----NGLDL-----GGKVISNV-----GKGTND 1614

Qy 1212 LYNKGMQYKGGYTKFDVSETDKGKESKVVKFRYFTNEG-----VMAKGVTVIDG--- 1262

Db 1615 TDAANVQQLNEVRNLLGLGNADQGNVNIADIKDPNCGSSSNRTVIKAGTVLGKGN 1674

Qy 1263 FTQYFEGEDGFQ-AKDCLVTFKGYFYFDAHTGNGIKDWT-----DGNANGDLSNVVKTQDKGSKKALLATYNAAG 1723

Db 1675 DTEKLATGVGVGVDK-----DGNANGDLSNVVKTQDKGSKKALLATYNAAG 1723

Qy 1307 WYFEDANGVAATGAQVINGQKLYFNEDGSO---VKG--GVKNADGTYSKYKEGFGELVT 1361

Db 1724 TNYVTNNPAEAI DRINEQGI RFFHVNDQNGPVPVQGRNGIDSSASGKHS-VAIGFQAKAD 1782

Qy 1362 NEFF-----TIDGNVWYVYAGANGKTVTGAQ-----VINGQHLVFNADGSOVKGVV 1407

Db 1783 GEAAVAIGRQFQAQNSQIAIGDNAQ-ATGDSQIAIGTGNVWAGKHSIGAIDPSTVK--- 1837

Qy 1408 KNADGTYG--KYNASTGERLTNEF-----FTTGDNNWYIIGANGKSVTGEVKIGDDTYFF 1460

Db 1838 --ADNSYSGVNNNQFTDATQTDVFGVGNITVTESNSVALGSN-----SAISAGTHAGTQ 1890

Qy 1461 AKDGKQVKGQTVSAG-NGRISYYGDSGKRAVS 1492

Db 1891 AKKSDGTAGTGTTTAGATGTVKFGAGTAVGAVS 1923

```
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-217

Query Match      3.8%; Score 301; DB 7; Length 726;
Best Local Similarity 31.1%; Pred. No. 3.6e-09;
Matches 68; Conservative 41; Mismatches 90; Indels 20; Gaps 5;

QY 1304 NGKWTYFPDANGVAATGAQVINGOKLYFENEDGSQVKGWVKNADGTSYKKEGFGELVTNE 1363
DB 157 NGWYILGADGRNVVTSHTIGKTYTFAQDGKQVKGAFADSDGNKHYYDRDSEGWYTR 216

QY 1364 FFTDGNVYIYAGAKTGTGAQVINGOHLYFNADGSQVKGWVKNADGTSYKYNASTGE 1423
DB 217 FVNDQGN-WYLLNNDGVPVVTGSIYVNGQSLYFNSDGSQVKGNEVEE-DGSLRYDYKNSGD 274

QY 1424 RLTFNEFTTCDNNWYIYGAKSVTGEVKGID-----DTYFFAKDGKQVKGQTVSAGNR 1478
DB 275 LLKRTSRITNGVNYQFDNDGNARAIDKIEVVVKTSLVVDSYEFGFSVSKI-----ILEFNHK 330

QY 1479 ISYVYDGSQ-----KRAVSTWIEIQPVVYVYFDKN 1508
DB 331 VTPAVVHAGAVVTTAGVQRKILNSVYVSNASHVYVYFDS 369

RESULT 10
US-11-013-759-11
; Sequence 11, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-11

Query Match      3.8%; Score 299; DB 7; Length 2314;
Best Local Similarity 19.5%; Pred. No. 2.4e-08;
Matches 374; Conservative 229; Mismatches 598; Indels 616; Gaps 93;

QY 11 KYVQHWTTIYASVALATVGLSVTTSVSADETDQ-----KTVTQSNSGTTAS 60
DB 239 KAEKG-TIAIGSNAQAINVGAL-----ALGADTRVDLDYGIALGYGSOILNNN----- 287

QY 61 LVTSPEATKADKRTWKADVLTPAKETNAVETATTTTQATAEAAATTATTADVAVA 120
DB 288 -----NNNNKAYV--PEGNGSNTKSKATGNGLFSGISSTIKRKIINVGA- 330

QY 121 PNKEAVVTTDAPVTTKEABEOPA-----TVKAEVNVTEVK---APEAALKDSEVEA----- 169
DB 331 -----GYEDTDVNVVQALKAVENLAKROIYFKGDDNGTGVKKLGETLTIKGETQADKLT 386

QY 170 -----ALSLKNIKNIDGKYVYVNEEDGSHKENFAITVNGQLLYFGKDGALTSS 216
DB 387 DNNNIGVWTDNNTGLKVKLAKNLSGLETVSTKNLTASEK--VTGVS-----GNNTAELQS 439

QY 217 STYSFTPGTNTI-----VDGFSINNRYASDSSEASFELIDGYLTADSWYRPASIIKD- 267
DB 440 GGLTFTP-TTNASTDKTYVGTDLKFT-----DNSNTALE-----DTTRITKDK 482
```

```
QY 268 -GVYWOASTAEDFRPLLMAMWPNVDQVNYLYNTMSKVFN-----LDKYSSTDKQETLKV 321
DB 483 IGFSNKAGTVDENKPYL-----DKDKLKVGNSLTNNNGGLTVANTTIGSINKQ--IQV 531

QY 322 AAKDIO---IKIBOKIOAEKSTOWLBRETISAFVKTOPQWNKETENTYKSGGGEHLOGGAL 378
DB 532 GADGIKFPADVNVNVVNAARFGTTRITEEIGFADADGKVKDKSPYLDK-----KQLQVGV 587

QY 379 LYVND-----SRTPWANS-----YRRLNRATATGCTIDK-SILDEQ----- 415
DB 588 KITKDSGINAGDKISNVVDAATDDTDAVYKQKQVQDADGALQSFSDINDEKQEFTIS 647

QY 416 -----SDPNHMGDFLLANDVDLSNPVVOAE---QLNQIHYLMWGSIVMGDKD--- 462
DB 648 NLYSNGNTENTPBTITFAGENGISISNDIAKGVKVGIDPGLNLTTPKLTGVSDDKDKTK 707

QY 463 -----ANFDGIRV----- 470
DB 708 LVIEQVASGNDTKNIIRGLSPTLPSITNAGVTRTEQGNITTSDEDEKSAASIGDILNTG 767

QY 471 -----DAVDNVDADMQLYTNV-----PREYVGVNKSEA-----NA 501
DB 768 FNLKNNNSVGVFSTVNTVDFIDGNATTAKVYVDETNOQTSKVTYDVNVDEKTIETLGDNG 827

QY 502 LAHISVLEASLNDNHYNK-----TDGAALAMENKQRLALIFSLAKPIKERTPAVSPL 555
DB 828 KTKNGKVTTLTTTNNANGKATNFSTTDNDAL-VNAKDIAENLNTLAKEL-HTKGTADT 885

QY 556 YNNTFNWTO---RDEKTDWINKDGSK-----AYNEGTV----- 586
DB 886 ALQTFKVKDGDATDDETIITVGKDGQNGKTVNTLKLKGENGTLVATNKGDTVFGINTQS 945

QY 587 -----KQSTIGK-----YNEKYGDASGNVVFIRAHDNVQ---DIAEIIKKEINP 629
DB 946 GLKAGDSITLNDKGLSINKNPASNEQIQVGADGVKFAKVDKNGSSTGIDGTSRTKQOI-- 1003

QY 630 KSDGFTITDAEMKQAFEIYNNKDMLS--SDKKYTLNNIPAAVYVLMQNMETITRVYVGDLYT 688
DB 1004 ---GFTGANGSLDTTKPHLTOKLKGVEVEIINTGINA-----CGKKITNIQSDDI-T 1052

QY 689 DDGHYNETKSPYYDTIVNLMSRIKYVSGGQORSYMLPTDGMNDMSDELVELVTRNBYTS 748
DB 1053 QNSNDAVTGGRVYDLKTEL-ESKIN-SAAKTAQNS-----LHEFSVA 1092

QY 749 VRYGKDIWNTANDTEGSKYRTSGQVTLVANN-----PKLNLDSAKLVNVEGKIH 798
DB 1093 DEQGNHFTVSNPYSSYDTSKTSVDITFAGENGITTKVKNGVVRVIGIDQTKGLTTPKLTV- 1151

QY 799 ANQYRALIVGTADG-----IKNFTSD--ADAIAAGVYKVEDNSGVLTFFG----- 841
DB 1152 GNNNGKIVIDSKDQNTITGLSNTLANVTNDGAGHALSQGLANDTDKTRAASIGDVINA 1211

QY 842 -----ANDIKGYETFDMSGFVAVVMPVVGASD-----NODIRVAPSTEAK 880
DB 1212 GFNLQNGEAVDFVSTYDVTDFIDGNATTAKVYDIDTSKTSKVYVDVNVNKNKTIEVTSK 1271

QY 881 KEG--ELTIKATEAVDSQLIEGFSNFQTIPOGSDPSVYVTRKIAENVLDLFSWGVTSPE 938
DB 1272 KLGVKTTTLTKTSANGNATKFS--ADGDALVKASDIATHLN-TLAGDIOQAK--GASQAS 1327

QY 939 MAPOFVSADGGTFL-DSVIQNGYAFADRYDLAMSNNKYGSKEDLRDALKALHKAIGIOAI 997
DB 1328 SSASVYDADGNKVIYDSTDCKYQVNDKQO--VDKN-----KEVAKDLVA-----QA- 1373

QY 998 ADMVPD-QIYQLPGKEVVTATRTDQAGRKIADAIIDHLSLYVANSKSGKDYQAKYGGEFL 1056
DB 1374 --QTFDGTLAQMWKSVINKEQVNDANKK--QGINEDNAFIKGLENAAKDTTKNAAVTV 1429

QY 1057 AELK--AKYPEPKVNNMISTGKPIDDSVKLQWKAEYFNGTNTV-ERGVGVYVLSDEATGK 1113
DB 1430 GDNLNVAQTPLTFFAGDTGTAKKGLGTLTIKGGQTD-----TNKLTDDNNIGVVAGTDG--- 1482
```

QY 1114 YFTVTKEGNFIPLOLTGKEKVTGPSS--DGKGIYFTGTS--TOAKSAFVTFNGNTYYDA 1171
Db 1483 -FTVK-----LAKDLNLNSVAGGTRIDKEGLISFVDANGQAKANTPVLISANG-----DL 1532
QY 1172 RGHMTNSEYSPNGKDVYRFLPNGLMISNAFYIDANGNTLYNSKGMYKGGYTKFDVSE 1231
Db 1533 GGRISNIGAAVDDNDVNFQKQNEVAKT-----VNLNNSQNSGASLSPVVD 1581
QY 1232 TD-----KOGKESKVVKFR-----YFTNEGVMAGVTVYIDGFTQYFGEDG--FOAKDKLV 1279
Db 1582 ANGKPINGTGRKPKQAKGADGKYHAN-----ANGVPV-----DKDGKPIITDADKLA 1629
QY 1280 TFKGTYYYDA--HT-----NGIKDTWNINGKYYFPDANGV----- 1315
Db 1630 NLAAGKPLDAGHQVVASLGGNSDAITLTNIKSTLPQIDTPNMGANAGQAQSLPSLSAA 1689
QY 1316 ----AATGAQVIN--GOKL-----YFNEDGSQVKGVVKNADGTYSKYK 1353
Db 1690 QQSNAASVKDLVNLGNFNLQTNHNOVDVFKAYDTVNFVNGTGADITS--VRSADGTMNIT 1747
QY 1354 EFGELVTVNEFFTDCNV-----WYVA-----GANGKTVTGAQVIN-- 1389
Db 1748 VNTALAAATD-----DGNVLKAKDGKFKYKADLLMPNGSLKAGKSASDAKTPTGLSLVNP 1803
QY 1390 -----GQHLNFADGSQVKGVVKNADGTYSKYNASTGERLTNEFFTTGDNWYIGAN 1443
Db 1804 AGKGSTGDAVALN---NLKAVFKSKDGTMTTTVSSDGISIQGK-----DNSSITLSKD 1854
QY 1444 GKSVTGEV-----KIGDDTYFFAKDGKOV----- 1467
Db 1855 GLNVGKVISNVGKGYKTDAAVQQLNEVRNLLGLGNAGNDN-----ADGNQVNIADIKK 1910
QY 1468 -----KGQTVSAGNRSIYYYGDSGKRAVSTWIRIQPVVYVYFDKNGLA 1511
Db 1911 DPNSGSSNRITKAGTVLGGKGN-----NDTEKLATG-----GVQGVVDKOGNA 1955

RESULT 11
US-11-052-554A-229
; Sequence 229, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 229
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-229

Query Match 3.5%; Score 280.5; DB 7; Length 619;
Best Local Similarity 27.8%; Pred. No. 3.8e-08;
Matches 81; Conservative 20; Mismatches 99; Indels 91; Gaps 10;
QY 1207 NGNTYLYNSKGOMYKGGYKTFDVSETDKDGKESKVVKFPYFTNEGVMAGVTVIDGFTQY 1266
Db 408 NGWTFYNTDGMATG-----LQNG-----SWYLYNSGMA-----TGWLOY 447
QY 1267 FGEDGFQAKDKLVTFKGTYYTDAHTGNGIKDT--WRNINRWYTFDANGVAATGAQVING 1325
Db 448 -----NGSWYLYNA---NGAMATGMAGVNSWYLYNANGAMATGMLOYNG 489
QY 1326 QKLYFNEDGSQVKGVVKNADGTYSKYKEGFGELVTNEFFTTGDNWYVYAGANGKITVCA 1385

Db 490 SWYLYNANGAMATGWAKVNGS-----WYLYNANGAMATGW 524
QY 1386 QVINGQHLNFADGSQVKGVVKNADGTYSKYNASTGERLTNEFFTTGDNWYVYTGANGK 1445
Db 525 LQYNGSWYLYNANGAMATG-----WAKVNGS-----WYLYNANGA 559
QY 1446 SVTGEVKIGDDTYFFAKDGKQVKGQTVSAGNRSIYYYGDSGKRAVSTWIE 1496
Db 560 MATGWVKDGTWYLYLEASGAMKASQWFKVSDK--WYVYNGLGALAVNTTVD 608
RESULT 12
US-11-052-554A-231
; Sequence 231, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 231
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-231
Query Match 3.5%; Score 276.5; DB 7; Length 701;
Best Local Similarity 29.3%; Pred. No. 7.4e-08;
Matches 78; Conservative 22; Mismatches 109; Indels 57; Gaps 7;
QY 1232 TDKDGKESKVVKFRYFTNEGVMAGVTVIDGFTQYFGEDGFOAKDKLVTFKGTYYTDAH 1291
Db 480 TPKTGMKQENGWYFYNTDGMATGLQNGNSWYLYNANGAMATGLQNG--NGSWYLYNA- 537
QY 1292 TONGIKDT--WRNINRWYTFDANGVAATGAQVINGQKLYFNEDGSQVKGVVKNADGTVS 1350
Db 538 --NGSMATGLQNGNSWYLYNANGAMATGLQYNGSWYLYNANGAMATGLQYNGS---- 591
QY 1351 KYKEGFGELVTNEFFTTGDNWYVYAGANGKTVTGAQVINGQHLVFNADGSQVKGVVKNA 1410
Db 592 -----WYLYNANGDMATGLQNGNSWYLYNANGDMATG----- 624
QY 1411 DGTYSKYNASTGERLTNEFFTTGDNWYVYIANGKSVTGEVKIGDDTYFFAKDGKQVKKQ 1470
Db 625 ---WLQYNGS-----WYLYNANGDMATGMVVKDGTWYLYLEASGAMKASQ 665
QY 1471 TVSAGNRSIYYYGDSGKRAVSTWIE 1496
Db 666 WFKVSDK--WYVYNGSGALAVNTTVD 689

RESULT 13
US-10-873-528-184
; Sequence 184, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787

; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 184
; LENGTH: 744
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-184

Query Match 3.4%; Score 267; DB 6; Length 744;
Best Local Similarity 27.1%; Pred. No. 2.7e-07;
Matches 73; Conservative 28; Mismatches 106; Indels 62; Gaps 6;
QY 1232 TDGDKESKVVKFRYFTNEGVMAKGVTVIDGFTQYFCEGDFQA-----KDKLVTFKGTYY 1287
DB 523 TPTGKQENGMYFYNTDGSMAIGMLQNNNGSWYLLNANGMATGWYKD-----GDTWY 576
QY 1288 FDAHTNGIKDTRWNINGKYYFDANGVAATGAQVINGQKLYFENEDGSQVKGWVKNADG 1347
DB 577 YLEASGAMKASQFKVSKWYVNSNGAMATGMLQYNGSWYLLNANGMATGMLQYNGS- 635
QY 1348 TSYKKEGFELVTNEFFTTDGNVYAGANGKTVTGAQVINGOHLVFNADGSQVKGWV 1407
DB 636 -----WYLLNANGMATGWAKVNGSWYLLNANGAMATG--- 668
QY 1408 KNADGTYSKYNASTGERLTNEFTTGDNNWYIYGANGKSVTGKVGKIGDDTYFFAKDGKQV 1467
DB 669 -----WAKVNGS-----WYLLNANGMATGWVXGDTWYILEASGAMK 706
QY 1468 KGTVSAGNRISYYGDSKRAVSTWIE 1496
DB 707 ASQWFKVSDK--WYVYNGLGALAVNTVD 733

RESULT 14
US-11-052-554A-260 | Application US/11052554A
; Sequence 260
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 260
; LENGTH: 1647
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-260

Query Match 3.2%; Score 255.5; DB 7; Length 1647;
Best Local Similarity 18.2%; Pred. No. 3.5e-06;
Matches 322; Conservative 222; Mismatches 607; Indels 621; Gaps 81;
QY 1 MENKILYKHKVKKQVWTVAVASVALATVGLSVTSSVSADETODKTVTOSNGTTAS 60
DB 1 MEKKQFSLRKYKSGTGFSLIGSVFL-----MMTTVAADE-----LS 38
QY 61 LVTSPEAT-----KEADKRTNTEADVLTPAKETNAVETATT---TNTQATAEAAATTATTA 113
DB 39 TWSEPTITNHTQOQHLNTNTESSAESQDTSQITPKTNREKEQPOGLVSEPTTELA 98

QY 114 DVAAVAVPNKEAVVTTDAPAVTTEKAEBOAPATVKAEEVNTVE-----KA 157
DB DTDAAPWAN-----TGPDATQKSASLP-----VNTDHDWVKTKGWDKYGKQG 144
QY 158 PEAAKDSVEEAALSILKNIDGKYYVNEGSHKE-----NFAITVNGQLLYFGKQGA 212
DB 145 KVVAVIDTGDPAHQSNRISDVSTAKVKSKEDMLAQKAAGINYGSWINDKVVVF----- 198
QY 213 LTSSTYSFTPGTTNIVDG--FSINNRAYDSSEASFELIDGYLTADSWYRPAS-----I 264
DB 199 -----AHNVYNSDNIKENQFDFEDWENFEFAEAPKAIKKHKIYRQSTQAPKETV 253
QY 265 IKDGVWQASTAEDPRPLMAMWPNVDQVNYLYNMYKVFNLDAKYSTSDKQETLKVAAK 324
DB 254 IK---TEETDGGSHDID-----WTQDDDTKYESHGMHVTGIVAGNS-----KEAATG 300
QY 325 DIQIKLEQKIQAEK--STOWLRETIISAFVKTQP-----QWNKETENYKGGGEDH 372
DB 301 FLGIAPEAQVMFMRVFPANDVMGSAESLFKATEDAVAGADVINLSLGTANGAQLSGSKP 360
QY 373 LQ-----GGALLYVNDSTRTPWA--NSDYRRRLNRATNOTGTIDKSI--- 411
DB 361 LMEATEKAKAGVSVVVAAGNERVYGSDDHDDPLAINDPDLVGLVCSPTGRTPTSVAAIN 420
QY 412 -----LDEQSDPNHMGDFFLANDVDSN--PVVQAEQLNQIHYLMNWSIVM 458
DB 421 WVIQRLMTVKELENRADLNHGA---IYSESVDKNIKDSLGYDKSHQPAY-----VK 470
QY 459 GDKDANFDGIRV-DAVNDVADMLQLY----- 484
DB 471 ESTDAGYKAOQVQKDKIALIERDPNKTYDEMIALAKKHGALGVLI FNNKPGQSNRMLTA 530
QY 485 -----TNYFREVYG-----VNKS-----EANALAHISLVEAWSL--- 513
DB 531 NCMGIPSAFISHEFGKMSQLNGNGTGSLEFDSVSVSKAPSKQGNENHFS---NWGLTSD 587
QY 514 -----NDNHYNDKTGGAALAMENKQRLALLFSIAKPIKERTPAVSP- 554
DB 588 GYLKPDITAPGGDIYSTYNDNHYGSOT--GTSMA---SPQIAGASLLVQVLEKTQPNLPK 643
QY 555 -----LYNYFTNTTQDEKTDWINKGSKAYNEDGTGVKQSTIGKY---NEKYGD 600
DB 644 EKIADIVKMLMSNAQIHVNPETKTTSPROQGAGLLNIDGAV---TSLYVTVGKDNYS 700
QY 601 AS--GNVVFIRADNNVQDII-----AETIKKEINPKSDGTTTDAEMK--OAFEI- 647
DB 701 ISLGNITDWTFDVTVHNLNSKDKTLRYDTELLTHVDFQKGRFTTSLSKTYQGEV 760
QY 648 -----YKNQML-----SSDKKYTLNIP--AAYAVML 672
DB 761 VPANGKVTVRVTMDVSQFTKELTKQMSNGYVLEGFVRPRDSQDDQLNRVNI PVGFKQF 820
QY 673 QNMETITR--VYVYDLYTDDGHYMETKSPYYDIIVNLMKSRISKVSGGQQRWLPDQK 731
DB 821 ENLVAEESIYRLKSQKGTGFYFDESQPKDDIYVG-----KHFTG-----LVTLGS 866
QY 732 MDNSVELYRTNEVTVSRY-----GKDIANTDTEGSKYSRTSGQVTVANNPKNLQDS 787
DB 867 ETNVSTKITSNDGLHLTGTFKQADGKFILEKN-AQGN-----PVLAISPGDNNQDFA 918
QY 788 AKLANYEMGKIHANQYRALIVGTAD-----GIKNFTSDADATAAGYVKET 832
DB 919 AFKGVFLRKY---QGLKASVYHASDKEHKNPLWVSPESFKGDKNFNSDIRFAKSTTLGT 975
QY 833 DNGVLTITFGANDIKYETFDMSGGFVAVVVPVVGASQNDIRVAPSTAEKKEGELT-----LK 888
DB 976 AFSGKSLTGAELPDGYHYVVSYY-----PDVVGAQRQEMTFDMILD 1017
QY 889 ATEAEDSOLLYBGFNS-FQTIIP-----DGSDSPVYTNRKIAE--NVYDLFKSNGVTSF 937
DB 1018 ROKPVLQATFPDETNRKPEPLKDRGLAGVRKDSVYFLERKDNKPYTIVINDSKYVSV 1077
QY 938 ENAPQFVSAD-DGTFDLSVIQNGYAFADRYDLAMSK---NNKYGSKEDLRDALKALHKAG 993

Db 1078 ENKTVRQADGSFI-----LPLDKAKLGDFFYYWEDF-----AG 1113
QY 994 IQAIADWVPDQIYQLPGKEVWVATRTDGAQRKIADAIIDHSLYVANSKSGKDYQAKYGG 1053
Db 1114 NVAIAK-LGDHLFQTLGKTPIKLKTG-----NYQTK--- 1145
QY 1054 EFLAELKAKYPENFKVNMIISTGKPIDDSVKLQWKAKAEYFNGTNVLERGVYVLSDEATGK 1113
Db 1146 -----ETLKDNLQEMTQ-----SD--TG- 1160
QY 1114 YFTVTKEGNFIPQLTGKVKITGPFSSDGKITGYSCTGTAQKSAFVTENG--NTYYFDA 1171
Db 1161 --LVTNQAQLAVVRNQPSQLTKMQD-----FFISPNEGDKDFVAFKGLKNNVNDL 1213
QY 1172 RGHMVTNSEY-----SPNGKDVYRFLPNGIMLSNIFY-IDANGNTLYNSKGQMYKG 1222
Db 1214 TVNVYAKDDHQKQTPWSSQAGASA-----SAIESTAWYGITARGS-----KVMPG 1259
QY 1223 GYTKPVSFETDKDGKESK-----VVKFRYFTNEGVMAGVTVIDGF-----TQ 1265
Db 1260 DY-QYVYVYRDEHGKHKQYTIISVNDKPKMITQGRFDTINGV-----DHFTPDKTK 1310
QY 1266 YFGEDGFQAKDLVTFPKGTYVYFDAHTGNGIKDWTWNINGKWYYPDANGVAATGAQVING 1325
Db 1311 ALGSSIVREEVFYLAKKNGRKFVTEGK-----DGI-----TVSD 1346
QY 1326 QKLYFNEDGSQVKGGVVKNADGTYSKYKEGFGELVTNEFFTTD---GNVWYVYAGANGKTV 1382
Db 1347 NKWY-----IPKNPDGYSYTIISKRD-GVTLSDYYLYVEDRAGNVSPATLRDLKAV 1394
QY 1383 -----TGAQVINGQHLVFNADGSQVKGGVVKNADGT-----YSKYNAST 1421
Db 1395 GKDKAVNFGLDLPVEDKQIVNFTYLVDRDADGKPIENLEYNNSGNSLILPYGKY---T 1451
QY 1422 GERLTNE-----FFTTGDNWYVI 1440
Db 1452 VELLTYDTNAKLESKIVSFSLADNFFQQV 1483

RESULT 15
US-10-873-528-185
; Sequence 185, Application US/10873528
; Publication No. US2005027681A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hanebro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 185
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-185

Query Match 3.2%; Score 251; DB 6; Length 693;
Best Local Similarity 29.4%; Pred. No. 1.8e-06;
Matches 63; Conservative 17; Mismatches 80; Indels 54; Gaps 5;
QY 1283 GKTYYFDAHTGNGIKDWTWNINGKWYYPDANGVAATGAQVINGQKLYFNEDGSQVKGGVV 1342
Db 522 GMYFY--NTDGSMTGWLQNGNSWYIYLNNGAMATGWLQNGNSWYIYLNANGSMATGWLQ 579

QY 1343 KQADGTYSKYKEGFGELVTNEFFTTDGNVWYVYAGANGKTVTGAQVINGQHLVFNADGSQV 1402
Db 580 NNGS-----WYIYNANGSMATGWLQYNGSWYIYNANGSMA 614
QY 1403 KGGVVKNADGTYSKYNASTGERLTNEFFTTGDNWYIIGANGKSVTGEVKIGDDTYFFFAK 1462
Db 615 TG-----WLQYNGS-----WYIYNANGDMATGWLKDGDTWYILEA 649
QY 1463 DGKQVKGQTVSAGNGRISYYGGDSGKRAVSTWIE 1496
Db 650 SGAMKASQWFKVSDK--WYIYVNGSGALAVNTTVD 681

Search completed: February 11, 2006, 20:59:01
Job time : 19.6383 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 31.3322 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-40
Perfect score: 7928
Sequence: 1 MENKIHYKLHKVKQWTTIA.....PGVYVFDKGLAYPRVLN 1518

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7928	100.0	1518	2 A44811	glucosyltransferas
2	3743	47.2	1599	2 S22737	glucosyltransferas
3	3515	44.3	1431	2 A45866	dextranucrase (EC
4	3296.5	41.6	1592	2 A38175	glucosyltransferas
5	3196.5	40.3	1475	2 B33135	gtfB protein precu
6	3155	39.8	1577	2 T30858	glucosyltransferas
7	3070	38.7	1375	2 JT0345	dextranucrase (EC
8	3067	38.7	1449	2 T30857	glucosyltransferas
9	3056	38.5	1449	2 T30552	glucosyltransferas
10	2804	35.4	1365	2 A41483	glucosyltransferas
11	2722	34.3	1290	2 JCS473	dextranucrase (EC
12	2638.5	33.3	1508	2 T31098	probable dextranu
13	582.5	7.3	2817	2 B97033	uncharacterized pr
14	522	6.6	2710	2 A37052	toxin A - Clostrid
15	419	5.3	2364	2 I40884	cytotoxin L - Clo
16	405	5.1	648	2 S10869	enterotoxin A - Cl
17	394	5.0	2178	2 S55805	alpha-toxin - Clo
18	367	4.6	2367	2 S70172	toxin B - Clostrid
19	366	4.6	2366	2 S10317	glucan-binding pro
20	344.5	4.3	563	2 A37184	probable peptidogl
21	319.5	4.0	1806	2 AF1717	cell wall-associat
22	300.5	3.8	2334	2 S32920	AAS surface protei
23	297	3.7	1463	2 T30290	hypothetical prote
24	294.5	3.7	1959	2 AG1085	cytotoxin RTX homo
25	285	3.6	1829	2 S35027	surface protein ps
26	280.5	3.5	619	2 A97887	hypothetical prote
27	280.5	3.5	619	2 A41971	cell wall-associat
28	278.5	3.5	1946	2 AE1449	
29	278.5	3.5	2167	2 AF1489	

30	277.5	3.5	1612	2 AB1347	probable peptidogl
31	276.5	3.5	701	2 H98120	choline binding pr
32	276	3.5	1983	2 G86643	hypothetical prote
33	275	3.5	1256	1 A43829	muramidase-release
34	270	3.4	1315	2 T28679	fibrinogen-binding
35	268.5	3.4	4199	2 S76412	hypothetical prote
36	268	3.4	1829	2 E81086	iron-regulated pro
37	267	3.4	744	2 F95013	pneumococcal surfa
38	266.5	3.4	1305	2 H41662	150K mating aggreg
39	265.5	3.3	1306	2 S22624	aggregation protei
40	265	3.3	1296	1 HMS01F	probable peptidogl
41	261	3.3	2044	2 AB1180	hypothetical prote
42	260.5	3.3	1385	2 D89824	hypothetical prote
43	258.5	3.3	1659	2 H97926	hypothetical prote
44	255	3.2	4152	2 T31102	filamentous hemagg
45	253.5	3.2	1993	2 AF1450	probable peptidogl

ALIGNMENTS

RESULT 1

A44811
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C>Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A44811; S22726; S28809
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J:Gen. Microbiol. 137, 2577-2593, 1991
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase ge
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GIP>
A:Cross-references: UNIPROT:Q00600; UNIPARC:UPI000000BEF31; EMBL:Z11873; NID:947526; PID:
C:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIP:81052)
C:Genetics:
A:Gene: gtfJ
C:Keywords: glucosyltransferase; hexosyltransferase
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match	100.0%	Score	7928	DB	2	Length	1518
Best Local Similarity	100.0%	Pred. No.	0				
Matches	1518	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MENKIHYKLHKVKQWTTIAVASVALATV	LGSLVTTSSVSADETQDKT	VTQSN	SGTTAS	60	
Db	1	MENKIHYKLHKVKQWTTIAVASVALATV	LGSLVTTSSVSADETQDKT	VTQSN	SGTTAS	60	
Qy	61	LVTSPKATKEADKRTNTEADVLT	TPAKETNAVETATTTNTQATAEA	TATTAD	VAAV	120	
Db	61	LVTSPKATKEADKRTNTEADVLT	TPAKETNAVETATTTNTQATAEA	TATTAD	VAAV	120	
Qy	121	PNKEAVTTDAPAVTTEKAEQ	PAVTVKAEVNVTEVKAP	EALKDSE	VEAALS	180	
Db	121	PNKEAVTTDAPAVTTEKAEQ	PAVTVKAEVNVTEVKAP	EALKDSE	VEAALS	180	
Qy	181	GKYYVNDGSHKENFAITV	NGQLLYFGKDGALT	TSSTYS	TFPGT	240	
Db	181	GKYYVNDGSHKENFAITV	NGQLLYFGKDGALT	TSSTYS	TFPGT	240	
Qy	241	SSEASFELIDGVLTD	ADSVYRPASIIKDGVTWQ	ASAE	DFRPL	300	
Db	241	SSEASFELIDGVLTD	ADSVYRPASIIKDGVTWQ	ASAE	DFRPL	300	
Qy	301	SKVFNLDAKYSSTDK	QETLKVAAKDIQ	IKIEQIK	QAEKST	360	
Db	301	SKVFNLDAKYSSTDK	QETLKVAAKDIQ	IKIEQIK	QAEKST	360	
Qy	361	TENYSGGGEDH	LOGGALLYVND	STPWANSY	RRLNRT	420	
Db	361	TENYSGGGEDH	LOGGALLYVND	STPWANSY	RRLNRT	420	

Db 511 EHNTAALSMONGLRSLIVHGLTRPVTKNGTGCARNASMKDLINGGYFGLSNRAEVTSY--- 567

Qy 575 DGSKAYNEDGTVKQSTIGKNEYKYGDSAGNYVFIIRAHNNVQDIIAEIIKKEINPKSDGF 634

Db 568 -----DQLGPAT--YLFVRAHDSEVQTVIADIISKIDPTDGF 604

Qy 635 TIIDAEMKQAFIYNKMDLSDDKYITLNNIIPAAAYAVMLQWMEITRYYVYGLDLYTDDGHY 694

Db 605 TFLDQLQKQAFIDYNADMLKVDKEYTHSNIIPAAVALMLQTMGAATRYVYGLDLYTDCNQYM 664

Qy 695 ETSPYVYDTIIVNLKMSIKYKVSQGAQORSYMLPTDGRMNSDVVELYRTNEVTVSVRGKD 754

Db 665 AKSPFDFQITLLKAPKPKVAGQTSYIHNLAGDGVSSAKD-----NKEVLSVRVQGD 719

Qy 755 IMTANDTEGSKYRTSQVTLVANNPKNLDSQAKLNVEMGKIHANQYRALIVGTADGI 814

Db 720 LMSKTDTEGGKYGRNSGMLTLIANNPDLKLADGETITVNMGAHQNAQYRPLLLGTEKI 779

Qy 815 KNFTSDADATAAGYVKETDSNGVLTFGANDIKGYETPDMGFGFVAVMVPVQASNDQDIRVA 874

Db 780 VSSLNDS--TKIVKYTDAQGNLVTADBEIKGFKTVDMSGYLSVMVPVGVATDDQNVLAK 836

Qy 875 PSTEAKKEGELTLKATEAYDSQLIYEGFSNFQITPDGSDPSVVTNRKIAENVDLFKSGV 934

Db 837 PSTRAYKEGDKYSSAALEAQVIYEGFSNFQFV--KEDSQTNKLJIAANADLFKSWG 894

Qy 935 TSFEMAPQFVSADDTGFLDSVIONGYAFADRYDLAMSNNKNGSKEDRLDALKALKHAGI 994

Db 895 TSFEIAPQYVSSKDGTFLDLSIENGVAFTDRYDPFMSKNNKNGSKEDRLDALKALKHKO 954

Qy 995 QATADWVPDQIYQIPGKEVVVATRTDGAGRKIADAIIDHSLYVANSKSGKOYQAKYGE 1054

Db 955 QVIADWVPDQLYPLGKEVVVATRTDTHGKVLDDTSLVNKLKYVTNTKSSGNDPQAOYGA 1014

Qy 1055 FLAELAKAPPEMFKVMNISTGKPIDDSVKLQKWAKEVNGTNNLERGVGVVLSDEATGK- 1113

Db 1015 FLDKLQKLPFIEFKEVMEASGKIDPSVKIQKWEAKYFNGTNIQKRGSDVVLSD---GKL 1071

Qy 1114 YFTVTRKGNFIPQLTGKEKVIITGFSDDGKGIYFTGTSQAKSAFVTNGNTYYPDARG 1173

Db 1072 YFTVNDKGFPLPAALTDGTKAKTGFAYDGTGVYIYTTSGTQAKSQFVYNGKQYFNDKG 1131

Qy 1174 HMVTNSEYSPNGKDVTRFPLNGIMLSNAFYIDANGNTYLYNSGQMYKGYTKPVDSETD 1233

Db 1132 YLVT-GEQITDGSN-YFFLPNGVWFMDGVRKNAKQSLVYKSGKLTQTGTGWK-EVTVKD 1188

Qy 1234 KDGKSKVKFRYFTNEGVMKAVTVIDGFTQYFEGEDGQAKDKLTFK----- 1282

Db 1189 DSGKEBF--YOYFFKGGIMATGLTEVEGKEKYFYDNGYQAKGVFPTKDGHLMPFCGDS 1246

Qy 1283 ----- 1282

Db 1247 GERKYSFPFEDGNWYYANDKGYVATGFTKVKGQNLVFNKGVQVKNRPFQVGDATYYAN 1306

Qy 1283 -----GKTYFFDAHTNGIKDTRWNINGK- 1306

Db 1307 NEGDLVLRGAQTINGDELDFDESGQVKBGFVNNPDGTTSYDAITGVKLVDSLVVDGQT 1366

Qy 1307 -----WTFPDANGVAATGAQVINGQKLYFNEGDSQVKGVV 1342

Db 1367 FNVDAKGVVTKAHTPGFYTTGDNWNWFYADSYGRNVTTGAQVINGQHLVFDANGQVKG 1426

Qy 1343 KNADGTSYKKEGFGELVNEPFTTDGNVYVYAGANGKVTGAQVINGQHLVFNADGSOV 1402

Db 1427 TWTDSRSFYHWTGDKLVSTFPATGHRVYADDRGNVVTGAQVINGQKLFDDTDGKV 1486

Qy 1403 KGVVKNADGTSYKYNASTGERJTNEFFTTGDNWNYIYGANGSKVTVGEVKIGDDTYFFAK 1462

Db 1487 KGAFATNANGRSYHYHNTGNKLVSTFFTSGDNNWYADAKGEVVVGEQTINGQHLVFDQ 1546

Qy 1463 DGKQVKGQTVSAGNRISYVYDGSGRKAVSTWIEIQPVYVFPDKNGLAY 1512

Db 1547 TGKQVKGATATNPDGSIYYDHTGEKAINRWVKIPSGQWYFNAQKG 1596

RESULT 3

A45866

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C;Species: Streptococcus mutans

C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004

C;Accession: A45866

R;Honda, O.; Kato, C.; Kuramitsu, H. K.

J. Gen. Microbiol. 136, 2059-2105, 1990

A;Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl

A;Reference number: A45866; MUID:91100958; PMID:2148600

A;Accession: A45866

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1431 <HON>

A;Cross-references: UNIPARC:UPI000017AC5C; GB:M29296

C;Keywords: glycosyltransferase; hexosyltransferase

F;181-201/Domain: cpl repeat homology <CP1>

F;1127-1146/Domain: cpl repeat homology <CP2>

F;1192-1211/Domain: cpl repeat homology <CP3>

F;1257-1276/Domain: cpl repeat homology <CP4>

F;1277-1297/Domain: cpl repeat homology <CP5>

F;1321-1340/Domain: cpl repeat homology <CP8>

F;1341-1361/Domain: cpl repeat homology <CP6>

F;1385-1404/Domain: cpl repeat homology <CP7>

Query Match 44.3%; Score 3515; DB 2; Length 1431;

Best Local Similarity 48.1%; Pred. No. 3.4e-156; Matches 724; Conservative 246; Mismatches 436; Indels 100; Gaps 31;

Qy 1 MENKIHYLKHVKKQWVTIIVASVALATVLGGLSVTSSVSAD---ETQKTVQ-SNSG 56

Db 1 MEIKRYKHVKKQWVTIIVASGLITL--GTTILGSSVSASETEQTSKVVTKSDD 57

Qy 57 TTASLVTSPKATKDRNTKEADVLTTPAKETNAVETATTTNTQATAEAA--ATTADV 115

Db 58 KAASESSQTDAPRTKQAQTEQTAQ-----SQANVADTSTSIKTPSQNITTTQANSDDK 112

Qy 116 AVAAVENEKAVVTDAPAVTTEKAEQAPATVRAEVNTEVKA-----PEAALKDSEVEA 169

Db 113 TVTNTKSEEAQITSEE-----RTKQSEEAQITASSQAL-TQAKAELTKOROTAAQENKNPVD 167

Qy 170 ALSLNKINIDGKYVYVNEDESGHKNFALTNGQLLYFGKD-GALTSSTSYSTPTGTNI 228

Db 168 LAALPNVKQIDGKYIYIGSDGQPKKNFALTNNKVLVFNKVTGALTDTSQYQKQSLTKL 227

Qy 229 VDGFSINNRAVDSSEASFELIDGYLTADSWYRPASIIKDGVTWQAETAEDFRPLLMWWP 288

Db 228 NNDYTPHNQIVNFPENTSLETIDNYVTADSWYRPKDIKNGKVTWASSSEDRLRPLLMSWP 287

Qy 289 NVDTQVNYLYNYSKV-FNLDAKYSSTDQKQETLKVAAKDQIKIEQKIQAEKSTQWLRETI 347

Db 288 DKQOTIAYLNYNNQOGLGTGENYTADSSQESLNLAQTQVVKIETKISQTTQQTQWLDRDI 347

Qy 348 SAFVKTPQPMWNETENYKSGGGEHLQGGALLYVNDSTRTPWANSDYRRLNRATNTGTI 407

Db 348 NSFVKTPQNNWNSQTSDESDTSAGEKDLQGGALLYNSDKTAYANSDFRLNRFTPTSGTK- 406

Qy 408 DKSILDEQSDPNHMGGFDFLLANDVLSNPVVAEQALNQIHYLMNMGSIIVMGDKDANFDG 467

Db 407 -----PKYPEDNSSGGYDFLLANDIDNSNPVVVAEQALNHLVLMNYSIVANDPEANFDG 461

Qy 468 IRVDAVDNVDADMQLQLYNTYFREYGVGNKSEANALAHISVLEAWSLNDHNDKDTGGAAL 527

Db 462 VRVDAVDNVDNADLLQIADSYLKAHYGVDKSEKNAIHNLSILEAWSNDPQYNKDTKGAQL 521

Qy 528 AMENKORALLFLSAKPIK-----ERTPAVSPLYNNTFNTTQORDEKTDWINKGSKAY 580

Db 522 PIDNKLRLSLLYALTRPLEKDNKNNEIRSGLEPVTISLN----- 562

Qy 581 NEDGTGVKQSTIGKNEYKYGDSAGNTVFIIRAHNNVQDIIAEIIKKEINPKSGFTITDAE 640

Db 563 -----NRSAGKNSER-----MANYIFIRAHSEVQTVIAKIIKAQINPKTDGLFTLDE 612
QY 641 MQQAEIYNKMLSSDKYTLNANIIPAAYAMVQNMMETITRVYVYGDLYTDDGHYMETKSPY 700
Db 613 LQAQFIYNEDMRQAKKYTQSNIPYAYALMSKDSITRLYVYGDYSDGQYMATKSPY 672
QY 701 YDTIVNLMKSRILKYVSGGAQARSYMLPTDQKWDNSDVELYRTNEVYTSVRYKDIWNTAD 760
Db 673 YDAIDTLKARIKYAAGGQDMKITVY--EGDSKSHMD---WDYTGVLVTSVRYGTGANEAYD 727
QY 761 TEGSKYSRTSGQVTLVANNPKMLQDSAKLVNEMGKIHANQYRALIVGTADGIKNFTSD 820
Db 728 -QGESEATKQGMVAVITNSNPSLKJLQNDKVIYNMGAHKNQBYRPLLLTTKQGLRSYTS- 785
QY 821 ADAIAAGYVYKETSNGVLFPFGANDIKGYETFFMSGEVAVVVPVGASDNODIKVAPSTEAK 880
Db 786 -DAAAKSLYRKNTDKGELVFDASDIQYLNPPQVSGYLAWVVPVGASDNQDVRVAASNKAN 844
QY 881 KEGELTLKATEAYDSOLIIYEGFSNFOI--IPDGSDFSVYTNRIENVDLFKSWGVTSPFM 939
Db 845 ATGOV--YESSALDSOLIIYEGFSNFODFVTKDSI---YTNKKIAQVQLFKSWGVTSPFM 900
QY 940 APOFVSADGTFLDSVIQNGYAFADRYDLAMSKNNKYGSKEDRLDAKALHAKGAGIQAIAD 999
Db 901 APOYVSSDGSFLDSIIQNGYAFEDRYDLAMSKNNKYGSQODMINAVKALHKSGLQVIAD 960
QY 1000 WYPDQIYQIPGKEVVTATFTDAGRKIADAIIDHLSYVANSKSGKDYQAKYGGSFABL 1059
Db 961 WYPDQIYVLPKGKEVVTATRVNDYGEYRKDSIKNTLYAANTKSGKDYQAKYGGSFABL 1020
QY 1060 KAKYPMFKVNMISGKPIDDSVKQWKAEPNGTINVLERGVGYYVLSDEATGKFTVTK 1119
Db 1021 AAKYPSIFNRTQISNGKKIDPSEKITAWKAKYFNGTINILGRGVGYVYLKDNASDKYFELKG 1080
QY 1120 EGNFIPLOLTGKEKIVTGFPSSDGKITYFGTSGTQAKSAFV--TFNGNTYFFDARGHMTN 1178
Db 1081 NQTYLPKQNTKE--ASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYFFDNGHMYVG 1139
QY 1179 SEYSPNGKDVYRFLPNGIMLSNIFYIDANGNTYLYNSKQMYKGGYTKFDFVSETDKDGE 1238
Db 1140 LQ-QLNG-EVQYFSLNGVQLRESFLENADGSKNYFGHLGNRYSGNYSPDNDS----- 1190
QY 1239 SKVVFRYFTNMGWAKGVTVIDGFTQYFGEQFQAKDKLVF--FKGKTYVFDHTGNGIK 1297
Db 1191 ----KWRYPFASGVMAVGLKTNNGTQYFDQDQYQVKGAWITGSDGKKGIFDGDGNGMAV 1246
QY 1298 DTWRN--INGKWTYFDANGVAATGAQVINGOKLYFNEGDSQVKGGVVKNADGTYSKYKEGF 1356
Db 1247 NRFANDKNGDWYLLNSDGLVGVQTINGKTYFFQDQKQIKGIITD--NGKLYFLANS 1305
QY 1357 BELVTNEFTTNGVNYXAGANGKTVTGAQVINGOHLVFNADGSOVKGGVVKNADGTYSK 1416
Db 1306 GELARNIPATDSQNNWYFSGDVAVTGSGTTIAGKKLYFASDGKQVKGFSV--TYNGKVHY 1364
QY 1417 YNASGERLTNEFTTNGDNNWYVIGANGKSVTGEVKGIDGTFFPAKDGKQVKGQTVSAGN 1476
Db 1365 YHADSGELQVNRFEADKQGNWYLLNSGEALTGSGQRINDQRFVFTREGKQVKGDVAYDER 1424
QY 1477 GRISYV 1482
Db 1425 GLIRYV 1430

RESULT 4
A38175
Glucosyltransferase precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004
C:Accession: A38175
J:AbO, H.; T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A:Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A:Reference number: A38175, MUID:91123227, PMID:1704006

A:Accession: A38175
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1592 <ABO>
A:Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:G217032; PIDN:BAAI4241.1; PID
F:1093-1112/Domain: cpl repeat homology <CP1>
F:1222-1241/Domain: cpl repeat homology <CP2>
F:1287-1306/Domain: cpl repeat homology <CP3>
F:1330-1351/Domain: cpl repeat homology <CP4>
F:1352-1371/Domain: cpl repeat homology <CP5>
F:1402-1420/Domain: cpl repeat homology <CP6>
F:1465-1484/Domain: cpl repeat homology <CP7>
F:1513-1532/Domain: cpl repeat homology <CP8>
Query Match 41.6%; Score 3296.5; DB 2; Length 1592;
Best Local Similarity 44.4%; Pred. No. 6.4e-146;
Matches 702; Conservative 240; Mismatches 463; Indels 175; Gaps 32;
QY 1 MENKIHYKHKVKKOMVTIAVASVA-LATVLGGLSVTTSSVSADETQDKVTVQSNSGTTA 59
Db 1 MEKNVRFKKHKVKKWVTLVSASATMLASALGASVASADTDITASDDSNQTVV----- 52
QY 60 SLVTSPEATKEADKRTNTKEADVLTPAKETNAVETATTTNTQATAEAAATATTADVAVA 119
Db 53 -----TCGQTTNNQATD-----QTSIAATAT-----SEQSASTDAATDQASAA 90
QY 120 VPKNEAVTVTADPAVTTKEABEQPATVKAENVN-----TEVKAPEALKDSEVEAAL 171
Db 91 EQTGTTASTDTAAQTNTTNAME-AKWVPTENENOGFTDEMLAEAKNVATAESDSIPSDLA 149
QY 172 SLKKNIKNDGKYYVNEGSHKENFAITVNGQLLYFGKGALTSSTSYTFTPGTTNIVDG 231
Db 150 KMSNVQKQVGGKYYVQDGNVKKNFVSVGDKIYFDETGAYKDTSKVDADKSSSAVSON 209
QY 232 ---FSINNRAYDSSSASFELIDGYLTADSWYRPASIIKDGVTQWQASTABDFRPLLMAMP 288
Db 210 ATIFAANNRAYSTSAKNFEAVDNYLTADSWYRPKSIILKDGKTWTESGKDDFRPLLMAMP 269
QY 289 NVDTQVNYLYNYSKVFNLDAKYSSTDQKQETLKVAAKDIQIKTEQKIQAKSTQWLRETS 348
Db 270 DTETKRYVYVNMKNVGVIGDKTYTAETSOADLTAAAEELVQARIEQKITSNNNTKWLREAS 329
QY 349 AFVKTQPOWNETEN--YKSGGEDHLOGGALLYVND--RTPWANSDYRRLNRTATNQTGT 406
Db 330 AFVKTQPOWNGESEKPY-----DDHLQNGALLFDNQTLTPTQSNRYLNKRTPTNQTS 384
QY 407 IDKSTILDEQSDPNHMGGFDFLLANDVDLSNPVYVQAEQLNQIHYLMNWSIVMGDKDANFD 466
Db 385 LQSRFTYNPNDP--LGGYDFLLANDVDNSNPVYVQAEQLNWLHYLLNFGSIYANDADANFD 442
QY 467 GIRVDVNDVADMLQLYTNYFREYGVNKSANALAHISVLBAWSLNDNHNNDKTDGAA 526
Db 443 SIRVDAEDNVADQLQISDSYLLKAAAYGIDKNNKANNHVSIVVEASDNDPTPLHDDGDNL 502
QY 527 LAWEKQRLALLFLSAKPIKERTPAVSPLYNNFTNTQDEKTDWINKDGSKAYNEDGTV 586
Db 503 MNNDKFRLLWSLWSLAKPTDVHRS--GLNPLHNSLVDREVDDR-----EVEIV 548
QY 587 KQSTIGKYNKYGDASGNVYFIRAHNNVQDIIIAEIIKKEINPKSDGFTITDAEMKQAFE 646
Db 549 P-----SYSFARHDSVEQDIIIRDIKAEINPNSFGYFTQEEIDQAFK 592
QY 647 IYNKMLSSDKKYTLNNIIPAAYAVMLQNMETTITRVYVYGDLYTDDGHYMETKSPYTIYN 706
Db 593 IYNEDLKSKDKKYTHYNVPLSYTLTLTKNGKSIIPRVYVYGDYFDDGQGMANKTNYVDAIES 652
QY 707 LMKSRIKYVSGGAQORSYWLPTDKMDNSDVELYRTNEVYTSVRYKDIWNTADTSGSKY 766
Db 653 LUKARKVYVAGGQAMQNVQI-----NGEILT SVRYKGLKQSD-KGDAT 697
QY 767 SRTSGQVTLVANNPKLNDQSAKLVNEMGKIHANQYRALIVGTADGIKNFTSDADAI 826
Db 698 TRTSGVGVVGMGNQPNFSLDGKV--VALNMGAHANQEYRALMYSTKQGVATYATDADASKA 756

Db	753	HSDQE--AAGLVRYTNRGELIFTAADIKGIANPQVSGYLGVWVPVGA--LIIKMFALR	807
QY	878	EAKKEGEL-TLKATEAYDSQLIYEGFSNFQTTIPDGDSPVYNNRKIAENVDLIFKSHGVTS	936
Db	808	LARPHQOMASVHQNAALDSRVMEPGFSNFQAF--ATKBEYTNVVIKAKNDKFAEIMGVTD	865
QY	937	FEMAPQVADGDTFLDSVIQNGYAFADRYDILAMSKNNKYGSKEDLRDALKALHKAGIOA	996
Db	866	FEMAPQVSDGSLFDSVIQNGYAFDRYDILAMSKNNKYGSKEDLRDALKALHKAGIOA	925
QY	997	IADWVDPQIYQLPGKEVVTATRTDAGAKRIADAIIDHSLYVANSKSGDYQAKYGGEF	1056
Db	926	MADWVDPQIYQLPGKEVVTATRTDAGAKRIADAIIDHSLYVANSKSGDYQAKYGGEF	985
QY	1057	AEIKAKYPEKFKNMISTGKPIDDSVKLQKWAEPNGTNVLERGVYVLSDEATGKYFT	1116
Db	986	EELQAKYPELFAKQISTGVPMDDPSVKIKQWAKYFNGTNILRGAGYVLKQATNTYFN	1045
QY	1117	VT--KEGNPIPLQTCGEKVIICFSSDGKGIYFGTSGTQAKSAFVTFNGNTYYPDARGH	1174
Db	1046	ISDNKEINFLPKTLNQLDSQV-GFSYDGGKGYVYISTSGYQAKNTFISEGDKWYFPDNGY	1104
QY	1175	MYTNSYSNPGKDVYFLPNIGMILSNAPFIDANGNTYLYNSKGOMYKGGYTKFDVSETDK	1234
Db	1105	MYTGAQ-SINGVNY-FLSNGQLRDAILKNEDETYAYTNGDGRYENGYYQF-----	1155
QY	1235	DGKESKVKFRYPTNEGVMAGVTVIDGFTQYFGEDGFOAKDLV-TFKGKYTYFDAHGTG	1293
Db	1156	--MSGV--WRHP-NGGMSVGLTVIDGQVQVFDGMYQAKGKFTVTADGKIRYFDKQSG	1209
QY	1294	NGIKDTW-RNINKWYYPDANGVAATGAQVINGOKLYFNEGDSGVKGGVVKNAADTYSKY	1352
Db	1210	NMYNRFIENEKGWLDGEDGAAVTGSGTQINGHLYFRANGVQVKGFEVTDHHRISYY	1269
QY	1353	KEGFGHLVNEFTTDCNWWYTAGANGKTVTAQVINGHLYFNADGSGVKGSGVKNADG	1412
Db	1270	DGNSGDQIRNFRVNAQOGWFFDNGGYAVTGARTINGQLLYFRANGVQVKGFEVTDHYG	1329
QY	1413	TVSKYNASTGERLTNEFTTGDNNWYIYGANGKSVTGKIGDVTYFFPAKQKQVGTQV	1472
Db	1330	RISYDGNSGDQIRNFRVNAQOGWFFDNGGYAVTGARTINGHLYFRANGVQVKGFEV	1389
QY	1473	SAGNGRISYYGDSGKRAVSTWIEIQPVYVYFDKNGLA	1511
Db	1390	TDHGRISYYDGNSGDQIRNFRVNAQOGWFFDNGGYA	1428
RESULT 6			
T30858			
Glucosyltransferase - Streptococcus salivarius			
C:Species: Streptococcus salivarius			
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T30858			
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.			
Infect. Immun. 63, 609-621, 1995			
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri			
A:Reference number: Z20909; MUID:95122197; PMID:7822030			
A:Accession: T30858			
A:Status: preliminary;			
A:Molecule type: DNA			
A:Residues: 1-1577 <SIM>			
A:Cross-references: UNIPROT:Q55245; UNIPARC:UPI0000088087; EMBL:L35928; NID:g662380; PID			
C:Genetics:			
A:Gene: gtfm			
Query Match			
Best Local Similarity 39.8%; Score 3155; DB 2; Length 1577;			
Matches 687; Conservative 243; Mismatches 494; Indels 194; Gaps 33;			
QY	1	MENKIHKLHKVKKQWTVIIVASVALATVILGGLSVTSSVSADETQDKT---VTQNSGT	57
Db	1	MENKVRFKLHKVKKQWTVIIVASVALATVILGGLSVTSSVSADETQDKT---VTQNSGT	57

QY	58	TASLVTSPEATKADKRTNKADVLTPAKETNAVETATTTNTQATAEAAT-----	108
Db	61	TASLVTTTVTTEQASQAQSVSAVATASVSHETSFQAATSAVSCQATAQAQTFSPVASQEVA	120
QY	109	-----TATTADVAAVAPNKEA-----	139
Db	121	VSSQTQSSQGEFTTTEQVTSQGTSTQVAGQTSQAQSTPSTVTEQARPRVLTNAAPAIATRAA	180
QY	140	EEQ-----PATVKAEV-----	159
Db	181	DSIRINARNRNTNITAGTTPNVTIITGPNTPKPNVTVTSPNGTRPNVTIIVTQPNQN	240
QY	160	AAALKDSEV-----EALSLK-----	205
Db	241	KPVQSPQSPQPNKPVQPNQPSLDYKPVASNLKTIQDKYIV-ENGVVKNAATELDGRLY	299
QY	206	YFGKDGALTSSTSYFTPGTTNIVDGFINNRAYDSSEASFELIDGILYLTADSWYRPASTI	265
Db	300	YFDETGAMVDQSKPIYRADAIIPNNSIYAVYQAYDTSSKSFHLDNFLTADSWYRPQIL	359
QY	266	KDGTWQASTABDFRPLMAMWPNVDQVNYLYNYSKVPNLDAKYSSTDKQETLKVAAKD	325
Db	360	KDGKWTASTEXDYRPLMTWMPDKVTQVNYLYNYSQQGFGNKTYTTDMMSYDLAAAAET	419
QY	326	IQIKIEQKIQAEKSTOMLRETISAFVKTPQPKNKETENYKGGEDHLOGGALLYVNDSP	385
Db	420	VORGIEERIGRGNTTWLQRLMSDFIKTPQGNWSESD-NLLVGGDKHLOGGALTFLNNSA	478
QY	386	TPWANSDYRRLNRTATNOTGT----IDKSILDEQSDPNHMGDFDILLANDVLSNPVQA	441
Db	479	TSHANSDFELMNRFTNTQGTGRKYHIDRS-----NGGYELLANDIDNSNPVQA	528
QY	442	EQLNQIHYLMNWSIVMGDKDANFDGIRVDAVDNVDADMQLQYTYNFRYYGVNKSSEANA	501
Db	529	EQLNWLHYIMNIGSILGNDPNSANFDGVRIDAVDNVDADILLOASDYFKEKYRVADNEANA	588
QY	502	LAHISVLPAWSLNDHNHNDKTDGAALAMENKORLALLSLAKPIKERTPAVSPLYNNTFN	561
Db	589	IAHLSILEAWSYNDHOYKNTKGAQLSDNPLRETLTTLTFLRKSNNR-GSLSERVITNSLN	647
QY	562	TTQRDEKTDWINKDGSKAYNEDGTVKQSTIGKYNKYGSDASGNVYFIRAHDDNNVDIAE	621
Db	648	NRSSE-----QKHTPRDANYIFVRAHDSFVQAVLAN	678
QY	622	IHKKSNPKSDGFTTIDAEKMQAPFYNKDMLSKSKYKTYTLNIPAAVAVMLQMETITRV	681
Db	679	IISKQINPKTDGFTTMDLQKQAFELYNADIAKADKXYTQYNIIPAAVATMLTKDSITRV	738
QY	682	YVGLYTDGHHVWETKSPYDITVNLKMSRIKYVSGQAQRSYWLPTDCKMDNSDVVELY	741
Db	739	YVGLYTDGHHVWETKSPYDITVNLKMSRIKYVSGQAQRSYWLPTDCKMDNSDVVELY	782
QY	742	TN--EYVTSVRYGKDIIMTANDTEGSKYSRTSQVTLVANNPKLNLDQSAKLNVEMGKIHA	799
Db	783	LNGYIIMSSVRYGKGAEEANQL-GTAETNQGMLVLTANRPMKLGANDRLVNMGAHK	841
QY	800	NQKYRALLIVGTADGINKNFTSDADAIAAGVYKETSNDNGVLTFGANDIKGVETEDMSGFVAV	859
Db	842	NQAYRPLLLSKSTGLATYLLKQSD-VPAGLVRYTQNGNLTFATDAGIAGHSTVEVSYLAV	900
QY	860	WYVPGASDNODIRVAPSTEAKEGELTLKATEAYDSOLIEYGFSPNFQTIIPDGDSPSVYTN	919
Db	901	WYVPGASDNODIRVAPSTEAKEGELTLKATEAYDSOLIEYGFSPNFQTIIPDGDSPSVYTN	956
QY	920	RKIAENVDLFXSWGVTSPMAPOFVSADDTGTFDLSVIQNGYAFADRYDILAMSKNNKYGSK	979
Db	957	RVIQAQNAKLFEKMGITSEFAPQYVSSQDGTFLDSIIENGYAFEDYDIIAMSKNNKYGSL	1016
QY	980	EDLRDALKALHKAGIOAIADWVDPQIYQLPGKEVVTATRTDAGAKRIADAIIDHSLYVAN	1039
Db	1017	KOLMDALRALHAEGISAIADWVDPQIYQLPGKEVVTATRTDAGAKRIADAIIDHSLYVAN	1076
QY	1040	SKSSGKDYQAKYGGEFLEBLKAKYPEMFKVNMISTGKPIDDSVKLQKWAEPNGTGNVLE	1099

```

Db 1077 TRTFGNDFOGKGGAELDELAKYPAIFERVOISNORKLUTNBSKITQWSAKYFNGSNIQG 1136
QY 1100 RGVGYVLSDEATGKYFYTVKEGFIPLQLTGKEKVTLPFSSDCGKGTTFGTSGTQAKSAF 1159
Db 1137 TGARYVLQDNATNQYFSVRAGOTFLPKQMT--EITSGFRFRVGDVQYLSIGYILAKNTF 1194
QY 1160 VTFNGNT-YFDDARGHMTNSEYSPNGKQVYRFLPNGLMISNAPFYIDANGNTLYNSKG- 1217
Db 1195 IQVGANQWYFYDFKNGNMTGEOVDGKK--YFELDNLGLRLHVLKQSDGHVYVYPKGV 1252
QY 1218 QMYKGGYTRFDVSETDKGKESKVKFRFYETNKGWAKGVTVVDIGTQYFGE--DGRQAKD 1276
Db 1253 QAENGFY-----DFAGPRQDV---RYPDGNGQMYRGLHDMYGTTFYFDEKTIQAKD 1301
QY 1277 KLVTFF-KGKTYYPDAHTGN-GIKDTRWNINGK-WYTFDANGVAATGAQVINGKLYFNEB 1333
Db 1302 KTRFADGRTRYPIPTGNLAVNRPQNPENKAWYLDNSGYAVTGLQTINGKQYFYDNE 1361
QY 1334 GSQVKGGVKNADGTYSKYKGEGELVTNEFFTTDGNVWYAGANGKTVTGAQVINGOHL 1393
Db 1362 GRQVKGHFVTNNQRY--FLDGSGBIAPSRFVTENKNWYVVDGNGKLVKGAQVINGNHV 1419
QY 1394 YFNADSGSVKGVVKNADGTYSKYNASGTGERLTNEFFTTDGNWYIANGKSVTGEVKI 1453
Db 1420 YFNNDYSQVKGAW--ANGRY--YDGDGQAVSNQFIQIAANQWYVLDNODGHKVTGLQNI 1474
QY 1454 GDDTYFFAKDGKQVKGQTSAGNGRISYYTGDSGKRAVSTWIRIQGVVYVYFDKGLA 1511
Db 1475 NKNVYFSGNGAQVKGKLLTV-QGKKCYFDAHTGTEQVNRVFEARAGCMYFNSAGQA 1531

RESULT 7
JT0345
dextranucrase [EC 2.4.1.5] precursor - Streptococcus mutans (strain GS-5)
N:Alternate names: sucrose 6-glucosyltransferase
C:Species: Streptococcus mutans
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C:Accession: JT0345; C33135
R:Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
Gene 69, 101-109, 1988
A:Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.
A:Reference number: JT0345; MUID:89137980; PMID:2976010
A:Accession: JT0345
A:Molecule type: DNA
A:Residues: 1-1375 <UED>
A:Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515
A:Experimental source: GS-5
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040565
A:Accession: C33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: UNIPARC:UPI000014E25D; GB:M17361
C:Genetics:
A:Gene: gtfC
C:function:
A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
C:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-1375/Product: glucosyltransferase #status predicted <MAT>
F:1126-1145/Domain: cpl repeat homology <CP1>
F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 38.7%; Score 3070; DB 2; Length 1375;
Best Local Similarity 46.2%; Pred. No. 1.9e-135;
Matches 661; Conservative 223; Mismatches 412; Indels 134; Gaps 33;

QY 1 MENKIYKHLHKQWVTIAVASALATVLGGLSVTTSSVSADETQDK--TVTQNSGTT 58

```

```

Db 1 MEKVRFKLRKVKRWVTVSIA--AVVTL---TSLGSLVKADSTDRQQAUTESQ----- 52
QY 59 ASLAVTSPEATKADKRTNKEDVLTTPAKETNAVETATTTTQATAEAATATTADAVVA 118
Db 53 ASLVTTSEAAKETLRTATDTSATSATSQPTATVTDNVSTTN-QST---NTTANTAFVVK 108
QY 119 AVPENKEAVVTTDA-PAVTTTEKAEQ-----PATVKAENVVTEVKAPE-- 159
Db 109 PTTTSEQATDSDKIITTSKAVNRLTATGKFPVANNNTAHPKVTVDKIVPIKPKGLK 168
QY 160 --AALKDSVEAALSJKNIDGKYYYNEDGSHKENFAITVNGQLLYFGKDGALTSS 217
Db 169 QPSSLSQDDIAALGNVKNIRKNGKYYKYKEDGTLQKNYALNMGKTFFFDETGALSNNT 228
QY 218 TYSFTPTGNT--LVGDFSNNRAYDSSEASFELIDGYLTADSWYRPASIIKQVGTWOAST 275
Db 229 LFSKGNITNDNTNSFAQYQNVYSTDVANFEHVDHLYTAESWYRPRIKDKKTTQST 288
QY 276 AEDFRPLLMAMPNVDTQVNLNMYSKVFNLDAKTSSTDKQETLKVAAKDIQIKIBQKIQ 335
Db 289 EXDFRPLMTWPDQETQRYVYNNQAOLGIHQTYNTATSPLOLNLAQTQTKIEBKIT 348
QY 336 AKSTQWLRETTISAFVKTPQPNKKTENYSGGGBDHLGGALLYYVNDNR-TPWANSDYR 394
Db 349 AKENTNMLRQTTISAFVKTSQAWNSDSEK---PFDDHLQKGALLYSNNKSLTQANSNYR 404
QY 395 RLNRATATNGTGTIDKSILDEQSDPNH-----MGGPDFLLANDVDLSNPVVAEQLNQH 449
Db 405 ILNRPTPTQTG-----KKDPRTADRTTGGVEFLLANDVDNSNPVVAEQLNMLHF 455
QY 450 LNNWGSIVMGDKDANPDGIRVDADVNDADMLQLVNTYPREYVGVNKSANALAHISVLE 509
Db 456 LNNFGNIYANDPDANFDSIRVDADVNDADLLQIAGDYLKAAKGIHKNDKAANDHLSILE 515
QY 510 AWSLNDNNHYNDKTDGAALAMENKQRLALLFSIAKPIKERTPAVSPLYNNTF-NTTORDEK 568
Db 516 ANSYNDTPYLHDDGDNMINMDNRLSLLSYSLAKPLNQRS-GWNPLITNSLVNRTDDNAE 574
QY 569 TDMWINKDGSKAYNEDGTQKSTIGKNEKYGDASGNVYFIRAHNNVDIIAIEIKKEIN 628
Db 575 T-----AAVPSYFIRAHNSEVQDLIRNIIRTEIN 604
QY 629 PKSDGFTTIDAEKQAFIYNKDMSSDKKYTLNNIIPAAVAVMLQNMETITRVYVYGLVLT 688
Db 605 FNVVGYSTTEIEKAPFIYNKDLATKEXKYTHYTALSYALLTNTKSSVPRVYVYGDMT 664
QY 689 DDGHYMETKSPYYDTIVNLMKSRIKYVSGGQARSYMLPTDGKMDNSDVELYRTNEVYTS 748
Db 665 DDGQYMAHKTINYEAETLLKARIKYVSGGQAMRNQ-----QVGN-----EIIITS 710
QY 749 VRYGKDIIMTANDTEGSKYSRTSGQVTLVANNPKMLDQSAKLVNEMGKIHANQYRALLIV 808
Db 711 VRYGKALKATDT-GDRTTTSQVAVIEGNPNLSRLKASDRVYVNVAMGAHKNQAYRPLLL 769
QY 809 GTADGINKFTSDADAIAAGYKVTDSNGVLTFPGANDIKGYETFDMSGFVAVVVPVGCASN 868
Db 770 TTDNGIKAYHSDOE--AAGLVRYTNDRGELIFTADIKGYANPQVSGYLGWVVPVGAAD 827
QY 869 QDIRVAPSTEAKKEGELTKATEYDSQLIYBGFNSFQTIIPDGSDPSVYTNRKIAENVDL 928
Db 828 QDVRVAASTAPSTDCK-SVHQNALDSRVWFEGFSNFQAF--ATKKEEYTNVVIKQVNDK 884
QY 929 FKSQVTSFEHAPQVVSADGTFDLSVTONGYAPADRYDLAMSKNKYGSKEDLDALKA 988
Db 885 FAEWGVTDFEMAPQVVSSTGDSFLDSVTQNGYAFDTRDYDLGSKENKYGKTADDLKA 944
QY 989 LHKAGIOAIAADWPDIYQOLPGKEVVVTRTDGACRKIADAIIDHSLVYVANSKSGKQY 1048
Db 945 LHSKGIKVMADWPDMYALPEKEVVVTRVVDKYGTPVAGSQIKNTLYVVDGKSGKQOQ 1004
QY 1049 AKYGEFELAEKAKYPEMFKNMISTGKPIDDSVKLGKQWKAERYFNKTVNLERGSGVYLS 1108

```

Db 1005 AXYGGAFLBELQAKYPELFARQKISTGVPMDFSVKIKQWSAKYFNGTNIILRGAGVYLKD 1064
Qy 1109 EATGKYFTVTKEGNFIPLOLT---GKEKVIITGFSDDGKIYTFGTSGTOAKSAFVTFNG 1164
Db 1065 QATNTYFSLVSDNTFLPKSLVNPNGHTSSSVTGLVPDGGYVYVYSTSGNQAQNAFISLGN 1124
Qy 1165 NTYYPDARGHMTNSSEYS PNGKDVIYRFLPGLMISNAFYIDANGNTYL--YNSKGOMYKG 1222
Db 1125 NMYFDNNGYMTGAO-SINGANY--FLSNGIOLRNAIY--DNGKNKVLSYGNDGRRYEN 1180
Qy 1223 GYTKFDVSETDKGSKSVKVKRYFTNEGVMAGKVTVIDGFTQYFGEDGFQAKDLV-TF 1281
Db 1181 GYILP-----GQQ-----WRYFON-GIMAVGLTRVHGAVQYFDASGFQAKQFIITA 1226
Qy 1282 KGTYYFDHAHTGNGIKDTH-RNINGKYYFDANGVAATGAOVINGOKLYFNEDEGSOVKGG 1340
Db 1227 DGKLRVDFDRDSGNQISNRNFRVNSKGEWFLFDHNGVAVTGTVTFNGQLYFKFNGVQAKGE 1286
Qy 1341 VVKNADGTYSKYKKGFGELVTFNEFFTDDGNVYVYAGAKTKVTGAQVING 1390
Db 1287 FIRDANGYLRYDDPNSGNEVRNFRVNSKGEWFLFDHNGIAVTGARVNG 1336
RESULT 8
T30857
Glucosyltransferase - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30857
R:Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
I:Infect. Immun. 63, 609-621, 1995
A:Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A:Reference number: Z20909; MUID:95122197; PMID:7822030
A:Accession: T30857
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1449 <SIM>
A:Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166B; EMBL:L35495; NID:g662378; PID
C:Genetics:
A:Gene: gtfl

Query Match 38.7%; Score 3067; DB 2; Length 1449;
Best Local Similarity 43.1%; Pred. No. 2.9e-135;
Matches 679; Conservative 204; Mismatches 495; Indels 198; Gaps 38;
Qy 1 MENKHYKLHKVKKQWVTIAVASVALATVLG-GLSVTSSVSADETQKTV----- 50
Db 1 MDKKVHKVHKVKKQWVTIAVGLSLGSAVSLSGTNDGVVQADEHTDATVAIPDITVDT 60
Qy 51 -----TQNSGTTASLVTSPKATKADKRNTEADVLTPAKET---NAVETATTTT 100
Db 61 GTVSNDDTTAAQDPTTAAVATNDVAT---DQATPTATFDLTDTTNTVAANAVIDTVATVGT 117
Qy 101 QATAEAATATTADVAAPVNPKEA VVTTDAP--AVTTEK-----AEEQP----- 143
Db 118 DRAATNDTATNDATVDTNNTTTDTTVDTRAATERRATGARGTGGRATPVNG 177
Qy 144 ----ATVKAENVNTEKAPKAEALKDSEVEAALS LKNIKNIDGKYVYVNEDESGHKNFAIT 199
Db 178 NTNANNTVTVNNDLPATNVNVDGP-----SHIKTINGQYVVEDDGTIRKNYVLE 230
Qy 200 VNGQLYFKED-GALTSSTSYF--TPGTTNIVDGFSIN-----NRAYDSSEASFELI 249
Db 231 RIGGSQYFNAETGELSNOKEYRFDKNGGTGSSADSTNTNVTNVNGDKNAFYGTGTDKIELV 290
Qy 250 DGLVTADSVYRPAS IIKDGVTAQASTAEDFRPLMAWPNVDVTQVNLNYSKV-FNLDA 308
Db 291 DGYFTANTYRPAEILKQKWEASTENDKRPULLTVWPSKAIQASLYNMKEQGLGTNQ 350
Qy 309 KYSSSTDQKFTLVAAKDIQIKIEQKIQABKSTOWLRETTISAFVKTOPWNKETENYSKGG 368
Db 351 TVTSPSSQTMQDAALVQKRIEERJAREGNTDWRATTIKNFVKIQPGMNSLENLD--- 407

Qy 369 GEDHLQGALLVYNDSTRPWANSYRRLNRTRATNOTGTIDKILDBOSDENHMGGFDFLL 428
Db 408 NNDHLQGALLVYNDSTRSHANSIRLLNRTPTSQTGKHNPXYTKDTSN----GAFEPFL 463
Qy 429 ANDVLSNPVQAEQLNQIHLYLMNMGSIUMGDKANFDGIRVDAVDNVDADMLQLYTNYF 488
Db 464 ANDINSNPAQAEQLNLHLYMNIIGTTTGGSEDEDFDGVRVDAVDNVNADLQLIASDYF 523
Qy 489 REYYGVNKEANALAHISVLEAWSLNDNHYNDKTPGAALAMENKORLALLFSLAKPIKER 548
Db 524 KAKYGADQSQAQAIKHLSTILEAWSHNDAYTNEDTKGAQLPMDPMPHLALVYSLLRPIGNR 583
Qy 549 TPVAPLNNNTPTTQRDEKTDWINKGSKA YNEDGTVKQSTIGKYNEKYGDASGNYVFI 608
Db 584 S-GVPEPLISNSLN-----DRSESGKNSKR-----MANYAFV 613
Qy 609 RAHDNNVQDIIAETIIEIKBINPKSDGFTITDAEMKQAFIYINKDMLSSDKKYTLNIPAA 668
Db 614 RAHDSVQSIIGQIIEIKNEINPQSTGNTFTLDEMKAFAIYINKDMSANKQYTOYNIPSA 673
Qy 669 AVLQNMETITRYYGDLTYTDDGHYMETKSPYYDTIVNLMKSRIKYVSGGQAQORSWLV-- 726
Db 674 ALMLTHKDTVRVYVYGDYTDGQYMAOKSPYDAIETLLKGRIRYAAAGQDMKVNYYIGY 733
Qy 727 -PTDGKMDNSDVELYRTNEVYTSVRYGKDIMTANDEGSKYRSRTSQVTLVANNPKLND 785
Db 734 GNTNG-----WDAAGVLTSVRYGTGANSASDT-GTAETRNQGMVIVSNQPALRV- 782
Qy 786 OSAKLNVBMGKIHANQYRALITVGTADGINKFTSDADAIAAGYVVKETDSNGVLTTCANDI 845
Db 783 -TSNLTINMGAHRNQAYRPLLLTTNDGVATVYLDSD--ANGIVKYTDGNGNLTFSANEI 839
Qy 846 KYETFDMSGFVAVWVPVGSADNQDIRVAPSTPEAKKEGELTLUKATEAVDSQLIYGFSGNF 905
Db 840 RGIENPQVDGYLAVWVPVGSASENQVRVAPSKKXSSG-LVYESAALDSQVIYEGFSNF 898
Qy 906 QTIPOGSDPSVYTNKIAENVDLFKSMGVTSPMAPQFVSADDTGFLDSVIQNGYAFADR 965
Db 899 QDFV--QNPQSYTNKIAENANLKFSGWITSEFAPQYVSSDDGSLDSVIQNGYAFDR 956
Qy 966 YDLAMSNNKYGSKEDLRDALKALHAKAGIQAATADWVPDQIYQLPGKEVVVTAATRDGAGRK 1025
Db 957 YDIGNSKNKYGLADLKAALKSLHAGISALADWVPDQIYNLPDGEVVVTAATRVNNGET 1016
Qy 1026 IADAIIDHSLYVANSKSGDYQAKYGGFEFLAELKAKYKPEMPKVMNISTGKPIDDSVKUK 1085
Db 1017 KDGAIIHSLYAAKTRTFGNDYQGYGAFLDELKRLYPOIFDRVQIISTGKMTTDEKIT 1076
Qy 1086 QWKAQYFNGTNNVLERGVGLSDEATGKYFTVTKEGNFPL-QLTCKEKVITGFSSDGGK 1144
Db 1077 QMSAKYMGNTNIDRGSEYVLKNGLNGYYGT---NGKRVSLPKVVGSNQSTNGDNQNGG 1133
Qy 1145 -----ITYFGTSGTQAKSAFVTFN-GNTYFFDARGHMTNSENSEYSPNGKDVYRFLP 1193
Db 1134 SKGFEKRLFSVRYRYNNGQYAKNAFIKNDGNYVYFDNSGRMAV-GEKTIQKQ-YFFELA 1191
Qy 1194 NGIMLSNAPYIDANGNTYLYNSKGOMYKGYTKFVDSSETDKDGKESKVVKRYFTNEGVM 1253
Db 1192 NGVQLRDGVRQNRQGVFYDQNGVLNANG-----KQDPKP----- 1227
Qy 1254 AKGVTVIDGFTQYFGEDGFQAKDLVTFKGKTYFYDAHTGNGIKDTRWNINGKYYVFDAN 1313
Db 1228 -----DNNNNASGRNQF-----VQIGNNV-----WAYYDGN 1253
Qy 1314 GVAATGAOVINGOKLYFNEDEGSOVKGVVKNADGTVSKYKEGFGELVTFNEFFTDDGNVWY 1373
Db 1254 GKRVTHQNINGQELFPDNNNGVQVAGRTV-NENGAI RYDYDANSGEWARRFAEIEPQVWA 1312
Qy 1374 YAGANGKTVTGAQVINGOHLYPNADGSQVKGGVKNADGTVSKYNASTGERLUTNEFFTGT 1433
Db 1313 YFNNDGTA VKGSONTINGDQLYFDQNGRQVKGALA-NVDGNLRYDYDVSNGELRYNRPHET- 1370
Qy 1434 DNNWYIYGANGKSVTGEVKIGDDTYFFAKDGGKQVKGQTVSAGNGRISYYIYDGSGBAVST 1493

Db 1371 DGSWYYFDGNGNAVKGMVNINGQLLFDNNGKQIKGHLVRV-NGWRYVFDPSGEMAVNR 1429

Qy 1494 WIEIQPGVYVYFDKNG 1509

Db 1430 WVEVSPGWWYVFDGEG 1445

RESULT 9

T30552
 glucosyltransferase N - Streptococcus salivarius (fragment)
 C:Species: Streptococcus salivarius
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T30552j
 R:Jaife, R.I.
 submitted to the EMBL Data Library, February 1998
 A:Description: Streptococcus salivarius V1477 gtfN.
 A:Reference number: 220854
 A:Accession: T30552
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-1449 <JAF>
 A:Cross-references: UNIPROT:O68542; UNIPARC:UPI00000B10FD; EMBL:AF049609; NID:G2935545;
 C:Genetics:
 A:Gene: gtfN

Qy	609	RAHDNNVDIIAEI KKEINPKSPGFTITDAEMKQAFIYNKMDLSSDKKYTLNIPAA	668
Db	614	RAHDSVQSIIQI KKEINPOSTGNTFTLDEMKAFIYNKMDRSANKQYQTQYNISAY	673
Qy	669	AVMLQNMETIRVYVYGDLYTDDGHYMETKSPYDITVNLKSRISKYVSGGAQARSYWL--	726
Db	674	ALMUTHKDTVPRVYVGYDMYTDDGYMAQSPYDAITFLKGRIRYAAAGQDMKNVYIGY	733
Qy	727	-PTDGKMDNSDELRYRTNEVYTSVRYGKDINTANDTEGSKYSRTSGQVTLVANNPKMLD	785
Db	734	GNTNG-----WDAAGVLTSVRYGTGANSASDT-GTAETRNQGMAVIVSNQALRL--	782
Qy	786	QSAKLNVEMKGIHANQKFRALIVGTADGIRKFTSDADAIAAGYKETSUNGVLTFGANDI	845
Db	783	-TSNLTTNMGAAHRNOAYRPLLLTTNDGVAIYFLNDS--ANGIKYTDGNGNLTFSA	839
Qy	846	KGYETFDMSGFVAVWVPVPGASDNODIRVAPSTEAKKEGELTLKATEAVDSOLIYEGESNF	905
Db	840	RGIRNPQVDGYLAVWVPVPGASENQDVRVAPSEKKNSSG-LVYESNAALDSQVIYEGESNF	898
Qy	906	QT-IPDGSDFSVYTRKIAENVDLFKSWGVTSEFMAQFVSADOGTFLDSVIQNGYAFAD	964
Db	899	QDFVP---NFSQYTNKKIAENANLFKSWGITSFEFAQYVYSSDDGSLDSVIQNGYAFD	955
Qy	965	RYDLAMSKNNKYSKEDLRDALKALHKAGIQAIADWPQDIQIYLPKGEVVTATRTDGAGR	1024
Db	956	RYDIGMSKDNKYGSLADLKAALKSLHAVGISAIADWPQDIQIYLPGEDEVVTATRVNNYGE	1015
Qy	1025	KIADAIIDHSILYVANSKSGKDYQAKYGGGFLAELKAKYEMFKVNMISTGKPIDDSVKL	1084
Db	1016	TKGGAIIIDHSILYAAKTRTFGNDYQKYGGAFLDELKRLYQIPIORVQISTGKRMVTD	1075
Qy	1085	KQKAEYFNGTINVLERGVGVYLSDEATGKFTVTKEGNFIPL-QLTGKEKVIITGFS	1143
Db	1076	TKWSAKMTNGTNILDRGSEYVLKXNGLNYGTYT---NGKVKVSLPKVVGSNOSTGNDQNGD	1132
Qy	1144	G-----ITYFGTSGTOAKSAFVTFN-GNTYYPDARGHMVTNSEYSPNGKDVYRPL	1192
Db	1133	GSQKFEKRLFPVRYRYNMGQYAKNAFTKDNDDGNVYYPDNSGRMAV-GEKTI	1190
Qy	1193	PNGIMLSNAPYVIDANGNTYLYNSKGQYKGGYTKFDVSEITDKGKESKVKVFRFTNEGV	1252
Db	1191	ANGVQLRDG-----YQNRARGQVF-----YYDQNGV	1216
Qy	1253	MAKGVTVIDGTQYFEGEDGFOAKDLVTFPKGKTYTYFDAHTGNGIKDTRNI-----	1307
Db	1217	LSA-----NGKQ---DPKPDNNNTSGRNQFVQIGNNVW	1247
Qy	1308	YYPDANGVAATGAQVINGOKLYNEDGSGVKGVGVKNADGTSYKYEKGFELVTNEPFTT	1367
Db	1248	AYYDNGKRYVIGHQNTINGQBELFPDNNGVQVKGRTV-NENGAIRYYDANS	1306
Qy	1368	DGNVWYYAGAKTIVTAQVINGOHLIYENADGSOVGGVVKNADGTSYKYNASTGERLTN	1427
Db	1307	EPGWAYFNNDDGTAVGSGQINIGODLYFDONGQVVGKALA-NVDGNLRYYYD	1365
Qy	1428	EFFTTGDNWYYITCANGKSVTGEVKIGDDTYFFPAKDGQVKQGTQVSAGNGRISYY	1487
Db	1366	RFHEI-DGSWYYPDNGNAVKGMVNINGQNLLFDNNGKQIKGHLVRV-NGVRY	1423
Qy	1488	KRAVSTWIEIQPGVYVYFDKNG	1509
Db	1424	EMAVNRWVEVSPGWYVYFDGEG	1445

RESULT 10

Accession: A41483
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Mar-2004
C/Species: Streptococcus sobrinus
C/Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
glycosyltransferase (EC 2.4.1.-) gtfs precursor - Streptococcus sobrinus
A41483

746 QY- KYTDSQGNLSFSAEQLSQVANAQVSGMIQVWVPVGAADNQDVRTSPSTQATKDCNI 804
 886 TLKATEAYDSQLIYEGFSNFQTIPDGSDPSVYTNRKIAENVDLFKSWGVTSSFEMAPQVFS 945
 805 -YHQS DALDSQVIYEGFSNFQAF--AQSPDYTNVIAKNGDLFKSWGITQFEMAPQVFS 861
 946 ADGTGFLDSVIQNGYAFADRYDLAMSKNNKYGSKEDLRDALKALHKAGIOAIADWVDDOI 1005
 862 SEDGTFLDSVILNGYAFSDRYDLAMSKNNKYGSKQDLANAIKGLQSAGIKVLSDLVENQL 921
 1006 YQLPGKEVVTATRTDGCAGRKIADAIIDHSILYVANSKSGKDYOAKYGGEEFLAEALKAKYPE 1065
 922 YNLPGKEVVTATRVNQYGQAKSGATNKTPYVANTRSYG--DYQEYGGKFLDDLQKLYPR 980
 1066 MFKVNMISTGKPIDDSVYKLGKQWKAERYFNGVTNVLERGVGYYLSDEATGKYFTVTVKEGNFIP 1125
 981 LFTSTKQISTGKPIDSPVKITNWSAKYFNGSNILGRGAKYVLSB--GNKYLNLADGKFLP 1038
 1126 LQIT---GKEKV-ITGESSDGKIGITFTGTSQTQAKSAFVTFPNGNTYTFDARGHMTVNSEY 1181
 1039 TVLNNYTGQPVSGANGFISKNGGIHLYLDKNGQEBVKNRFEISGSWYTFDSDGRMATGK-- 1096
 1182 SPNGKDYVPELPMGLMISNAFYIDANGNTVLYNSKGMQYKG--GVTFPDVSETDKDGKESK 1240
 1097 TKTGNDTYLFMPNGKQLKEGVWYDGK-KAYYDDNGRTWTKGFVEFRV-----DGDQ-- 1148
 1241 VKFRFYTNBEGVMAKVTVIDGFTQYFGEDGFOAKDLVTFKGTYYTFDAHTNGIKDT- 1299
 1149 --KWRIFNGDGTIAIGVSLDNRFLFDAGYGVQVKGQTVTINGKSYTFDADQGDVQTDN 1206
 1300 -----WRNNGKYYFDANGVAATGAQVINGQKLYFNEDSGSVKGGVKNADGTY 1349
 1207 ANPAPQOAGWKLQDNQWYRKDGQLLTGEQITDGQVFPDQNGVQVKGGTATDASGVL 1266
 1350 SKYKEGFELVTNEFTTGDGNVWYAGAKCTVTGQVINGOHLNFNADSGOVKGGVVKV 1409
 1267 RFYDRDQGHQVKGWYSTDNDNVVYNSGQVLTGLTIDGQTVYFDDKGIQAKGKAVMD 1326
 1410 ADGTYSKYNASTGERLTNEFTTGDNNWYVIGANG 1444
 1327 ENGNLRYFDADSGNMLRDRWKNV-DGNWYTFNENG 1360
 RESULT 11
 JC5473
 C:Species: Leuconostoc mesenteroides
 C:Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
 C:Accession: JC5473
 R:Monchois, V.; Willemot, R.M.; Remaud-Slimeon, M.; Croux, C.; Monsan, P.
 Gene 182, 23-32, 1996
 A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from
 A:Reference number: JC5473; MUID:97136686; PMID:8982063
 A:Accession: JC5473
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1290 <MON>
 A:Cross-references: UNIPROT:Q48756; UNIPARC:UPI000017ABDF; GB:U38181
 C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose
 C:Genetics:
 A:Gene: dsrA
 C:Keywords: glycosyltransferase; hexosyltransferase
 P:78-870/Domain: catalytic #status predicted <CAT>
 F:922-1290/Domain: glucan-binding #status predicted <GCB>
 Query Match 34.3%; Score 2722; DB 2; Length 1290;
 Best Local Similarity 42.9%; Pred. No. 3.1e-119;
 Matches 582; Conservative 215; Mismatches 427; Indels 132; Gaps 28;
 QY 173 LKNIKIDGKYYVNDGSKHENFAITVNGQLLYFGKD-CALYSSSYSTFTPGTTNIVDG 231
 DB 1 MPNVKHDGKVPYGDGQPKKPTTLIDKPYFDKDTGALSND-----KQYVSEL 53

QY 403 QTG-----TTDKSILDSQDPNEMHGDFLLANDVDLSNPVQABQLNQIHYLMWGSIVM 458
DB 471 QTGVPRYTTIDQS-----KGGPELLANDVNSPNVQSEQLNMLHYLMFNGSITA 520
QY 459 GKDNANFDGIRVDADVNDADMLQLVTNYFVGVNKSEANALAHISVLEAWSLNDNHY 518
DB 521 NDSANFDGIRVDADVNDADLLQTAADYFKAAYGVKNDATANQHLSLIEDWSHNDPPY 580
QY 519 NDKTGAALAMENKQRLALLFSLAKP IKERTTPAVSPLYNNTFTTTQRDEKTDWINKDGSK 578
DB 581 VKDFGNQLTMDDYVHMTQLWSLTKDMRMR-----GTMRFDYDYLVR----- 624
QY 579 AYNEDETQVQSTIGKNEKYGDASGYVFIHNDNNVQDIIAEIIEIKKENPK-SDGFTTIT 637
DB 625 --NHDST--ENT-----AIPNYSFVRAHDSVQTVIAQII--SELHPDVQNSLAPT 669
QY 638 DAEMKQAFIYNKMLSSDKKYTLNNI PAAYAVMLQNMETITRVVYGDLYTDDGHYMETK 697
DB 670 ADQLAEAFKIYNDEKQADKKYQYNMPSAYAWMLLTNKDTPRVYGGDLYTDDGQTMANK 729
QY 698 SPYDYITVNLMSRIKYVSGQAQRSYMPLPTDGKMDNSDVELYRTNEVYTSVRYGKDINT 757
DB 730 SPYFDAINGLLKSRIKYVAGQS-----MAVD-----QNDILTNVRYGKGAMS 772
QY 758 ANDTEGSKYSRSTSGQVTLVANNPKMLDOSAKLVNEMGKIHANQKVRALIVGTADIKNF 817
DB 773 VTDS--GNADTRTQIGIVIVSNKENLALKSGDVTTLHMGAAHKQNAFRLLTGTTANLSYY 831
QY 818 TSDADAIAGYKETEIDNSGVLPFGANDIKGYETFDMSGFFVAVVVPVPGASDNODIRVAPT 877
DB 832 DND-----NAPKYNTDQGLLIPDNTEIYGVNRNPQVSGFLAVVVPVPGADSHQDARTLSD 886
QY 878 EAKGEGELTKATEAYDSOLIEYEGFSNFQTIIPGSDPSVYTNRKIAENVNDFKSMGVTSF 937
DB 887 TAHHDGK-TFHSNAALDSQVIEGFSNFQAPATNIED--YTNVIAKNQQLFKWGIITSF 943
QY 938 ENAPQPVASDDGTFLDSVTONGYAFADRYDLAWSKNNKYGSKEDIRDALKALHKAGIOAI 997
DB 944 OLAPQVRSSTDTSFLDSIIQNGYAFETDRYDLGVTGTPTKYGTVDQLRDAIKALHANGIOAI 1003
QY 998 ADWVPDQIYOLPCKEVVTRATDAGAKTADALIDHSLVYVANSKSGKDYQAKYGERLPA 1057
DB 1004 ADWVPDQIYNLPQGEIATVTRINSYGDKTNSDIDQSLYVIOSRGGGK-YQAQYGAFUS 1062
QY 1058 ELKAKYPEMFKVNMISTGPIDDSVKLQWKAEYFNGTNVLRGVGVYLSDBATGKYFTV 1117
DB 1063 DIQKYPALFETKQISTGLPMDPSQKITEMSGKYFNGSNIQKGAGYVLKDSGTDQYKYV 1122
QY 1118 TKEGN---PIPLQTSKEKVIITGFPSSDGKIITYFGTSGTOAKSAFV-TPNGNTYYPDARG 1173
DB 1123 TSNNNNRDELPQLT--DDLSETGCFVRDNIQWVYVYTLGSLYLARNFTIQDNGNYYYPDSTG 1181
QY 1174 HMVTNSEYSPNGKDVYRPLNGIMLSNAFYVIDANGNTLYVNSKGQYKGYTKFDVSETD 1233
DB 1182 HLVTGFGQNNHH--YFFLPNGIELVQSEFLQNAAGSTIYFDQKGRQVFNQIY-----TD 1233
QY 1234 XDKGESKVVKRYFTNEGVM-AGKVTVIDGFTQYFGEDGFQAKDLVT-FKGKTYVYFDAH 1291
DB 1234 QTGTA-----YFQNDGTMTVTSGFTEIDGHKQYFYKNGTVQKGVPSDTHGVFLEA- 1286
QY 1292 TGNIGKDTWR---NINGKWYFDFANGVAATGAQVINGOKLYRNEDGSQVKG----- 1339
DB 1287 -GNGNVATORFAONSQOGWFLYLGNDGIALTGLQTINGVQNVFYADGHQSGKGFITIQNHV 1345
QY 1340 -----GUVKNADGTSKYKEGFGELVTNEFFTTDGNVWVYAGAKTVTGAQV 1387
DB 1346 LYNPLTGAITTGMQIQIGKIF--VPDNTGNMLTNGYYQTLQGWLUHLSTQGPADTGLVN 1403
QY 1388 INGQHLIYFNADGSQVGVVKNADGTYSKY-NASTGERLTNBEFTTGDNNWYIYGANGKS 1446
DB 1404 INGNLXYFQANGRQVKGQVFTDPIITNVSYMTNATDGSVAFNDYFTY-QGQWLYLTDNSYQL 1462

QY 1447 VTGEVKIGDDTYFFAK-DGKOVKGQTVSAGNGRISYYYGDSG 1487
DB 1463 VKGFKVNNKLOHFEITGVQTKSAHIIVNNR--TYIFDDQG 1502
RESULT 13
B97033
uncharacterized protein, related to enterotoxins of other Clostridiales [imported] - Cl
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: B97033
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97033
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2817 <KUR>
A:Cross-references: UNIPROT:Q97K42; UNIPARC:UPI00000CA0A0; GB:AB001437; PIDN:AAK79053.1.
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1079
Query Match 7.3%; Score 582.5; DB 2; Length 2817;
Best Local Similarity 22.3%; Pred. No. 8.3e-19;
Matches 382; Conservative 179; Mismatches 533; Indels 621; Gaps 96;
QY 36 TTSSVSADETQDKVTYQNSGTTASLVTSPKATKADKRTNTKEADVLTPAKETNAVEFA 95
DB 153 STSSAQTKGSNDNNIPSNNTSTNTSKNENPSNTD-----IKTEPANAPIKD----- 200
QY 96 TTTNTQATAEAATATTADVAVAAPNKEAVVTT---DAPAVTTKEABEQPATVKAENV 152
DB 201 -TPNQSDSALAKNKALSNLNLADSSQTSKYVSSNNDAPKVTNTTSTDK-----KASNLN 254
QY 153 TEVKAPEALKDSEVEAALSUKNIKNIDG-KYYVNEGSHKENFAITVNGQLLFGKDG 211
DB 255 ND-----SODGWYTK-----DGKKYYV--GVQKGFQ-SINKSIYYFNDDG 294
QY 212 ALT-----SSSTYSGTPTGTTNIVDGFSINNRAYDSSASPELIDGILTADSWYRPASII 265
DB 295 SMQGTWLKYNSSYYFDASGVMLTGLQNINGTYGYNDDG-KLLTGLQAINNNYYFV-- 351
QY 266 KDGVTWQASTAEDFRPELLMAWPNVDYTVNYLN---YMSKVFNLDKYSSTDKQETLKVA 322
DB 352 NDGV-----MQTGMITCNDISKYYPDNGNGVMOTGLVHNNKYVFGNDGKLLTG 399
QY 323 AKDIOIKIEQKQAEKSTQWLRETISAFVKTOPQNKETENYKGGEDHLOGGALLYVN 382
DB 400 LQNI-----NNYT-----YYF 410
QY 383 DS---RTPWANSD---YRRLNRTATQGTGIDKSIDLEQSDPNHMGDFLLANDVDL 434
DB 411 DSNQVMDTWITIDGSKYFVSVN--GVMTGII-----YISYYYGPFANDGKL 456
QY 435 SN--PVQAE-----QLNQIHYLMNMGSI-----VMGDKDANPD-----GTR 469
DB 457 LTGLQVINGNSYYPDTNGIRLVSRMITIDGKYDFYFNQDGLILTNINYGKYYFYISGVK 516
QY 470 VDAVDNVADMLQLVTNTPREY-----GVNKSEA-----NALAHISVLEAW-SL 513
DB 517 QTGLQNIIDG-----NYY--YFSSSGIMQTKIDGKTYFYFGDNGIRQI-----GWITY 563
QY 514 NDNHYNDKTDGAALAMENKQRLALLFSLAKPIKERTPAVSPLYNNTF----- 560
DB 564 QNNKYFNSDG-----SMQTDUKIYSYSTSP--YNTHYYGFDNDGKLLIT 607
QY 561 -----NTTQRD-----EKTDWINKSGKAYNEDGTVKQSTIGKNEYKYGDSAGNYVFTIR 609
DB 608 GLQTIKNGTYYPDSNGISQMGWVNIIDGKDFYFNSNSIMTNWVNIINDEXY-----YFYI- 660

QY 610 AHDNNVQDIIAEIKKEINPKSDGFTITDAEMKQAFIYNKMDLSSDKKTYTLNIPAYA 669
Db 661 ---NNV-----KOTGFQYINGK-----YYFDPD-----G 682
QY 670 VMLQNMETIT-RVYGYDLXTDDGHYMETKSPYDITIUNLMSRKIKYVSGAQRSLWLP 728
Db 683 IMGTGFTISGNYILD---DNG-----VKGTGWTI---KGKDYFDGNGWIMNYW-- 729
QY 729 DGKMDSDVELYRTNEVYTSV-----RYGKDINTANDTEGSKYSRTSGQVTLVANNPKL 782
Db 730 ---FDNDKTYTYINGNMOTGAISINNHYGFD---DNGIMQTGWQRINGRTTYFDNN-- 780
QY 783 NLQDSAK--LNVEMGKHANQKRALIVGTADGINKFTSDADALAAGYVKETDSNGVLTP 840
Db 781 ---GAATGLTVTEGKTYFYNTYAYL-----DTGFIYF 811
QY 841 GANDIKGYEFDMSGFVAV--WVPVGAS---DNODIRVAPSTEAKEGELTLKATEAYDS 895
Db 812 NNN-----YYFLDNGGVRTGWINYSNNRYLSDTVRV----- 845
QY 896 QLIYEGFSNFQITPDGSDPSVYTNRKIAENVDLFKSWGVTSEFMAPOFVSADDTGLDSV 955
Db 846 -----TGFTI-DG-----NKYIFDSSGA---MCTSFITV----- 871
QY 956 IQNGYAFADRYDLAMSKNNKYGSKEDLRDLKALHKAGIOAIADWVPDQIYQLPGKEVVT 1015
Db 872 --NG-----NTYGFSDK-----GI-----MLTGWQTIL 892
QY 1016 ATRTDGAGRKIADAIIDHSLVANSKSS-----GKDY--QAKYGGEFLLA--ELKAK 1062
Db 893 SSNYS-----SYNIYFNSDGSQAQGFYTLGKTYFEPNYGMYMLGNYINGK 941
QY 1063 YPEMFKNMISTGKPIDDSVKLK-----QMKAEYFNGTNVLERGVGVVLS 1108
Db 942 YYYFDNDGVIQTGWITDRSKYYLDPGSAVTFQFNINGKYYFNSGIMQTLGVYVND 1001
QY 1109 EATGKYFTVTKENFIPLQLTGKEK---ITGFSSDGK-----ITYFGTS---GTOAKSA 1158
Db 1002 -----YGFDDNGHI---LTGMHSINGVIYIYFDSTGKAQGFVTLGKTYFYNTNMTG 1052
QY 1159 FVTFNQNTYFEDARGHMT-----NSE---YSPNGKDV-----YRFLPNGIMLSN 1200
Db 1053 FVNANNLYFDEGVMQGWIMNYSNRYFSAFGASVTGFTQIDGNKYCFDSNG----- 1107
QY 1201 AFYIDA---NGNTLYNSKGM-----YKGGYTK-----FDVSETDKGKSKVKV 1243
Db 1108 AIYTDVVTINGSTYGFNTGIMLTGHTIRYNGYSSYFNTYFNSDGTAKTGFFYLNK 1167
QY 1244 FRYFT-NEGVMAKGVTVIDGFTQYFGEDGFQ-----AKDKLVTFKG 1283
Db 1168 TTYFNPDSGRMLQGYQYINGNHYYFAPDGTMTGWTNGSSKYLLDPGAAVTLGLQTNG 1227
QY 1284 KTYYPFAH---TCNGIKDTRWNINGKWWYFDANGVAATGAOVINGOKLYFNEGDSQVKG 1340
Db 1228 NKYCFDSNGILOHNGI-----FYIGNTYGSDNNGIMLTGLQLINGLYLFCFNSDGS-VK 1282
QY 1341 VVKNADGTY---SKYKEGFGELVTNE-FFTIDGNV--W-----YYAGANGKTVTGAQ 1386
Db 1283 LVTYLGKTYFYDSYVSFGFNINNTYFNGDGTMTGWNYGYYYLYLNDGSIKVTGQ 1342
QY 1387 VINGOHLVFNADGSOVKGGVKNADQTSKYNASTGERLTNEFFTTGDNNTYIYGANGKS 1446
Db 1343 TIDGNKYFYDYGA--KTGIV-NIDGNYYGFN-NSGVMLTGMQHINGST--YFNSNGIA 1396
QY 1447 VTGEVKIGDVTFFAKDGK-QVKGQTVSAGNRISYYIGDSKRAVST-----WTE 1496
Db 1397 NTGFTIYLGKTYFYDSYGRMOIGSMTI--NG-TSYFYFANGVMKMTSTDSPNTLAVGWVR 1452
QY 1497 -----IQPGVYVYFDKNGL 1510
Db 1453 DSYYYQYLYNAAAGTKLTGLQTDGNTY-YFDSNGI 1486

RESULT 14

A37052
toxin A - Clostridium difficile
C;Species: Clostridium difficile
C;Date: 31-Jan-1992 #sequence revision 31-Jan-1992 #text_change 24-Sep-1999
C;Accession: A37052; A60991; S21897; S2437; S08638
R;Dove, C.H.; Wang, S.Z.; Price, S.B.; Phelps, C.J.; Lyster, D.M.; Wilkins, T.D.; Johns
Infect. Immun. 58, 480-488, 1990
A;Title: Molecular characterization of the Clostridium difficile toxin A gene.
A;Reference number: A37052; MUID:90129305; PMID:2105276
A;Accession: A37052
A;Molecule type: DNA
A;Residues: 1-2710 <DOV>
A;Cross-references: UNIPARC:UPI000016EAD4; GB:M30307; NID:g144925; PIDN:AAA23283.1; PID
R;Wren, B.W.; Clayton, C.L.; Tabagchali, S.
FEMS Microbiol. Lett. 70, 1-6, 1990
A;Title: Nucleotide sequence of Clostridium difficile toxin A gene fragment and detecti.
A;Reference number: A60991
A;Accession: A60991
A;Molecule type: DNA
A;Residues: 1-154 <BIC>
A;Cross-references: UNIPARC:UPI0000178891; EMBL:X60984
R;von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerbo
Mol. Gen. Genet. 233, 260-268, 1992
A;Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A;Reference number: S2434; MUID:92293124; PMID:1603068
A;Accession: S2437
A;Molecule type: DNA
A;Residues: 1-92 <VON>
A;Cross-references: UNIPARC:UPI0000178892; EMBL:X60984
A;Note: the four fragments shown in reference A60991 correspond to four types of repeat
d with repeats ordered ABCDCDABCDABCCDABCCDABCCDABCDABCDABCD
Nucleic Acids Res. 18, 1629-1630, 1990
R;Sauerborn, M.; von Eichel-Streiber, C.
A;Title: Nucleotide sequence of Clostridium difficile toxin A.
A;Reference number: S08637; MUID:90221894; PMID:2109310
A;Accession: S08638
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-55, 'V', 57-2079, 'L', 2081-2549, 'S', 2551-2710 <SAU>
A;Cross-references: UNIPARC:UPI0000178893; EMBL:X51797
C;Genetics:
A;Gene: toxA
C;Superfamily: Clostridium difficile toxin A; cpl repeat homology
C;Keywords: cytotoxin; enterotoxin
F;1820-1839/Domain: cpl repeat homology <CP01>
F;1840-1860/Domain: cpl repeat homology <CP02>
F;1861-1881/Domain: cpl repeat homology <CP03>
F;1933-1952/Domain: cpl repeat homology <CP04>
F;1953-1973/Domain: cpl repeat homology <CP05>
F;1974-1994/Domain: cpl repeat homology <CP06>
F;1995-2015/Domain: cpl repeat homology <CP07>
F;2067-2086/Domain: cpl repeat homology <CP08>
F;2087-2107/Domain: cpl repeat homology <CP09>
F;2108-2128/Domain: cpl repeat homology <CP10>
F;2129-2149/Domain: cpl repeat homology <CP11>
F;2201-2220/Domain: cpl repeat homology <CP12>
F;2221-2241/Domain: cpl repeat homology <CP13>
F;2242-2261/Domain: cpl repeat homology <CP14>
F;2315-2334/Domain: cpl repeat homology <CP15>
F;2335-2355/Domain: cpl repeat homology <CP16>
F;2356-2376/Domain: cpl repeat homology <CP17>
F;2377-2397/Domain: cpl repeat homology <CP18>
F;2449-2468/Domain: cpl repeat homology <CP19>
F;2469-2489/Domain: cpl repeat homology <CP20>

Db 732 -----QDSITVSANOYEVRIINEBGRILD---HSGRWINKESIIKD-----ISSK 775
Qy 277 E--DPRPLMAMPNVNDTOVNYLNYMSKV--NLDAKYSTTDKQOETLKVAARDIQIKIEQK 333
Db 776 EYISFNPK-----ENKIIVKSYLHELSTLLQEIERNANSSDIDLEKKVMLTECEINVASN 831
Qy 334 IQAEKSTOWLEITISAFVKTOPOWNETENYSKGGEDHLOGGALLYNDSTRPFWANSY 393
Db 832 ID-----RQIVEGRIEAKNLTSDSINYIK-----NEF 859
Qy 394 RRLNRTATNOTGIDKISILDEQDPNHMGFD---FLANDV-----DLSN 436
Db 860 KLI-----ESISDSLYDLKHONGLDUSHFISFEDISKTENGFRIRINKETGN 907
Qy 437 PV-VQAEQLNQIHYLMNWSIVMGDKDANPDG-----IRVDAVDNVDADMQLQYTNV 487
Db 908 SIFETEKEITFSEYATHISKEISNIKOTIPDNVNGKLVKKVNLDAHAHEVNTLNSAFFIOS 967
Qy 488 FREYGVNKESEANALAHISVLEAWSLNDNHNDKTDGAALA-----ME 530
Db 968 LIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKYVELVSTALDETIDLLPTLSE 1027
Qy 531 NKQRLALLF---SLAKPIKERTPAVSPLYNNTENTTORDEKTDWINKGSKAYNEDGTVK 587
Db 1028 GLPIIATIIDGSLGAAIKELSETNDPLLRQEIEA-----KIGIMAVN-----LT 1072
Qy 588 QSTIGKYNEKYDASGNVYF-----IRAHDNN---VODIIABIIKKEINPKSDGF 634
Db 1073 AASTAIWTSALGIASGILLVPLAGISAGIPSLVNNELLIQDKATKI-----DYF 1124
Qy 635 T-ITDAEMQAPFIYNKMDLSSDKYTLNIPPAAYAVMLQNMETIT-----RVY-----YG 684
Db 1125 KHISLAETGAFTLLDDKIIMPODDLVLSEIDF-----NNSITLGRKEIWRABEGSG 1177
Qy 685 DLVTDG-GHMETKS-----PY---YDTIVNLMKSRIKY-----VSGGQAQRSY-----W 725
Db 1178 HLTDDIDIDHFFSPSYTRKFWLSIYD-VLNIKKEIDFSKOLMVLDPNAPNRVFGYEMGW 1236
Qy 726 LPTDGRMDNSVELY-RTNEVY-----TSVRYGKD-----754
Db 1237 TPGFRSLDNDGTKLLDIRDHYEGQFYRYFAPIADALITKLPKYVEDINVRILNDGNTR 1296
Qy 755 -----IMTANDTEGSKYS--RTSGQVTLVANNPKLNLDQS-----787
Db 1297 SFIVPVITTEQIRKRLNSYFYSGSGSYSLSPYNNIDILNLVENDTWIDVDNVVKNIT 1356
Qy 788 -----AKLNVE-----MGKIHANOKYBALIVGTADGINKFT 818
Db 1357 IESDEIQGELIENILSKLNIEDNKIILNNHTINFYGDINESNRFISLTFISILEDI-NII 1415
Qy 819 SDADAIAGYVKETDSNGV-LTFGANDIKGYETFDMSGFVA---VMVPVGASDNODIRVA 874
Db 1416 IEIDLVSYSKYILLSCNCKLIENSIDIQ--QKIDHIFNGEHQKYPYISYIDNETKYNG 1473
Qy 875 PSTEAKEGELTLKATEAVDSQLIYEGFSNFOTIPDGSPSVYTNR-----KIAENV-- 927
Db 1474 FIDYSKKEGLFTAERFSNESIIRNIY-----MPDSNNLFYSSKDLDIRIINKGDVK 1525
Qy 928 -LPSKNGVTSFEMAPQFVSAD-----DGTFLDSVIQNGYAFADRYDLAMSKNKGYSKE 980
Db 1526 LLIGYFFDKMDKVSLSFTIEDTNTIKLNGVYLD---ENGVA-----QILKFMNNAKSALN 1577
Qy 981 DLRLDALKALHKAQIOAI--ADWVPQOIYOLPGKEVVVTRTDGAGRKIAD-----1028
Db 1578 TSNLNMNFIENIKNIFFYNLDPNIEFILDNTNFIISGNSISIQELICDCKKNTPQYFI 1637
Qy 1029 -----AIIDHSLYVANSKS-----SGK-----DYQAK--YGEFELAEKAKYPE 1065
Db 1638 NEFIKETSYTLVGNRQNLIVEPSVHLDSDGNSISSTVINFSQKLYGIDRYVKNVLIAPN 1697
Qy 1066 MF--KVNMLSTGKPIDDSVKLQWKAAYEN---GTNVLERGCVYLSDEATG-----1112
Db 1698 LYTDEINITPVYKPNYICEVILDANYINEKINVINNDLSIRYVWNDNDGSLILIANSE 1757

Qy 1113 -----KYFTVTKEG-----NFIPLOLTGKEKVTGFS-----SDG-----1142
Db 1758 EDNPQVKIRFVNVFSDTAADKLSFNSDKQDVSVSKIIISTSLAAYSDFDPDYBGLV 1817
Qy 1143 -----KGITYFOTSGTOAK-----SAPVTNGNTIYFD-ARGHMVTN 1178
Db 1818 SLNDYFYINSFGNMVSGLIYIINDSLYFPPKNNLITGTTIDGKNKYFDPDKSGAASI 1877
Qy 1179 SEYSPNGKDVY-----RPLPNGIMLSN-----AF 1202
Db 1878 GEITIDGKDYFNNKQILQVGVINTSDGLKYFAPAGTLDENLEGESVNFIGKLNIDGKIY 1937
Qy 1203 YIDAN-----GNTYLYNSK-GOMYKG-----GYTK--PDVSETDKDGKESKVVKF 1244
Db 1938 YPEDNRAAAVEWKLDDETTYFNPKNKTEALKGLHQIGDNKYYPDDNGIMQGTITINDKV 1997
Qy 1245 RYFTNEGVMAKGVTVIDGFTQYFGEDE-----FOAKDK-----1277
Db 1998 FYFNNDGVNQVGYIEVNGKYFYFGKNGERQLGVFNTPDGFKFPKDDDLGTEEGELTY 2057
Qy 1278 --LVTFKGTIYFD---AHTNGIKDTRWNINGKWWYFDAN-GVAATGAQVINGOKLYF 1330
Db 2058 NGLNFKGIYFDDISNTAVGWGTLDD---DGSTIYFDDNRAEACIGLTVINDCKYF 2112
Qy 1331 NEDGSQVKGGVKNADGTYSKYKEGFGELVNTNEFFTDDGNVWYVYAGANGKTVTGAOVING 1390
Db 2113 DDNGIRQLG-----FITINDNIFYFS-ESGKIELGYQNING 2147
Qy 1391 QHLYFNADGSOVKGGVKNADGTYSKYNASTGERLTNEFFTDDGNWYVYIANGKSV--T 1448
Db 2148 NYFYIDESG-LVLIGVFDTPDG--YKYFAPLN-----TVNDNIY-----GOAVKVS 2190
Qy 1449 GEVKIGDDTYFPAKDGKQVKGOTVSAGNCRISYVYGDSCBRAVSTWIEIOPGVVYVFDKN 1508
Db 2191 GLVRNEDVYF---GETYKIETGIENETDKYFDPETKKAYK-GINVVDDIKIYF DEN 2246
Qy 1509 GL 1510
Db 2247 GI 2248

Search completed: February 11, 2006, 19:42:57
Job time : 44.3322 secs

THIS PAGE BLANK (USPTO)

RESULT 1	Q00600_STRSL	Q00600_STRSL PRELIMINARY;	PRT; 1518 AA.
ID	Q00600_STRSL	Q00600_STRSL PRELIMINARY;	PRT; 1518 AA.
AC	Q00600;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-FEB-2005 (TrEMBLrel. 29, Last annotation update)		
DE	Glucosyltransferase-I.		
DE	Name=gtfI;		
OS	Streptococcus salivarius.		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=1304;		
OX	[1]		
RN	NUCLEOTIDE SEQUENCE.		
RP	STRAIN=ATCC 25975;		
RC	STRAIN=ATCC 25975;		
RC	MEDLINE=92148377; PubMed=1838391;		
RA	Giffard P.W., Simpson C.L., Milward C.P., Jacques N.A.;		
RA	"Molecular characterization of a cluster of at least tow		
RT	glucosyltransferase genes in Streptococcus salivarius ATCC 25975."		
RT	J. Gen. Microbiol. 137:2577-2593(1991).		
RL	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=ATCC 25975;		
RC	Jacques N.;		
RA	Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.		
RL	EMBL; Z11873; CAA77900.1; -; Genomic DNA.		
DR	EMBL; M64111; AAZ26896.1; -; Genomic DNA.		
DR	PIR; A44811; A4811.		
DR	HSP; P06653; 1GVM.		
DR	GO; GO:0016740; F:transferase activity; IEA.		
DR	GO; GO:0009250; P:glucan biosynthesis; IEA.		
DR	InterPro; IPR002479; CW binding.		
DR	InterPro; IPR003318; Glyco_hydro_70.		
DR	Pfam; PF01473; G binding_1; 2.		
DR	Pfam; PF02324; Glyco_hydro_70; 1.		
KW	Transferase.		
SEQ	SEQUENCE 1518 AA; 167730 MW; DAA41F17098B59A CRC64;		
Query Match	100.0%; Score 7928; DB 2; Length 1518;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 1518; Conservative	0; Mismatches 0; Indels 0; Gap		
QY	1 MENKHYKLHKVKQWVTIAVASVALATVLGGLSVTTSSVSADETQDKVTQNSGTT		
DB	1 MENKHYKLHKVKQWVTIAVASVALATVLGGLSVTTSSVSADETQDKVTQNSGTT		
QY	61 LVTSPEATKEADKRTNKEADVLTPAKETNAVETATTTNTQATAEAATATTATTADVA		
DB	61 LVTSPEATKEADKRTNKEADVLTPAKETNAVETATTTNTQATAEAATATTATTADVA		
QY	121 PNKEAVTTDDAPVTEKAEQPAVKAEVVNTEVKAPAEALKDSEVEAALSUKNIK		
DB	121 PNKEAVTTDDAPVTEKAEQPAVKAEVVNTEVKAPAEALKDSEVEAALSUKNIK		

QY 181 GKYYVNEBDSGHENFAITVNGQLYFGKDGALTSSTYSFTPGTTNIVDGSINNAYD 240
DB 181 GKYYVNEBDSGHENFAITVNGQLYFGKDGALTSSTYSFTPGTTNIVDGSINNAYD 240
QY 241 SSEASFELIDGVLTDADSWRPAIIKDGVTWQASTAEDFRPLLMAWPNVDQVNLNLM 300
DB 241 SSEASFELIDGVLTDADSWRPAIIKDGVTWQASTAEDFRPLLMAWPNVDQVNLNLM 300
QY 301 SKVFNLDAKYSSTDQETLKVAAKDIQIEQIOAEKSTOMLRETISAFVKTQOWNKE 360
DB 301 SKVFNLDAKYSSTDQETLKVAAKDIQIEQIOAEKSTOMLRETISAFVKTQOWNKE 360
QY 361 TENYKGGGDEHLOGGALLVYNDSTRPWANSYRRLNRATNTQTGTIDKSLDEQSDPNH 420
DB 361 TENYKGGGDEHLOGGALLVYNDSTRPWANSYRRLNRATNTQTGTIDKSLDEQSDPNH 420
QY 421 MGGFDFLLANDVLDLNPVVOAQLNOIHVLMWGSTVMGDKDANFDGIRVDADVNDADM 480
DB 421 MGGFDFLLANDVLDLNPVVOAQLNOIHVLMWGSTVMGDKDANFDGIRVDADVNDADM 480
QY 481 LQLYTNYPREYGVNKSSEANALAHISVLEAWSINDHNNDKTDGAALAMENKQRLALLFS 540
DB 481 LQLYTNYPREYGVNKSSEANALAHISVLEAWSINDHNNDKTDGAALAMENKQRLALLFS 540
QY 541 LAKPIKERTPAVSPLYNNTFNTQREKTDWINKDGS KAYNEDGTGVTQKSTIGKYNEKYGD 600
DB 541 LAKPIKERTPAVSPLYNNTFNTQREKTDWINKDGS KAYNEDGTGVTQKSTIGKYNEKYGD 600
QY 601 ASGNVYFIRAHNNQDIIAEIIEKEINPKSGDFTTDAEMKOAEFEIYNKDWMLSDPKYT 660
DB 601 ASGNVYFIRAHNNQDIIAEIIEKEINPKSGDFTTDAEMKOAEFEIYNKDWMLSDPKYT 660
QY 661 LNNIPAAVAVMLQNNMETIRVYVYGDLYTDDGHYMETKSPYYDTIYNLMKSRIKYVSGQA 720
DB 661 LNNIPAAVAVMLQNNMETIRVYVYGDLYTDDGHYMETKSPYYDTIYNLMKSRIKYVSGQA 720
QY 721 QRSYMLPTDGKNDNSVELYRNEVYTSVRYKODIMTANDTEGSKYSRSTSGQVTLVANNP 780
DB 721 QRSYMLPTDGKNDNSVELYRNEVYTSVRYKODIMTANDTEGSKYSRSTSGQVTLVANNP 780
QY 781 KLNLDQSAKLVNEMGKIHANOKYRALIVGTADGINKFTSDADAIAAGVYKETSNGVLTF 840
DB 781 KLNLDQSAKLVNEMGKIHANOKYRALIVGTADGINKFTSDADAIAAGVYKETSNGVLTF 840
QY 841 GANDIKGYETFDMSGFVAVVVPVGSADNODIRVAPSTEAKKEGELTLKATEAYDSOLIYE 900
DB 841 GANDIKGYETFDMSGFVAVVVPVGSADNODIRVAPSTEAKKEGELTLKATEAYDSOLIYE 900
QY 901 GFSNFQTTIPDGSDPSVYTRKIAENVDLPKSGVGTSPFEMAPQFVSADGTFDLSVIQNGY 960
DB 901 GFSNFQTTIPDGSDPSVYTRKIAENVDLPKSGVGTSPFEMAPQFVSADGTFDLSVIQNGY 960
QY 961 AFADRYDLAWSKNNYKSGKEDRDALKALHKAGIOAIADWVPDQIYQLPKGKVVATATRD 1020
DB 961 AFADRYDLAWSKNNYKSGKEDRDALKALHKAGIOAIADWVPDQIYQLPKGKVVATATRD 1020
QY 1021 GAGRIADAIIDHSLYLVANSKSGDYQAKYGEFLAELKAKYPEMFKNMISTGKPIDD 1080
DB 1021 GAGRIADAIIDHSLYLVANSKSGDYQAKYGEFLAELKAKYPEMFKNMISTGKPIDD 1080
QY 1081 SVKLQWKABYFNGTNVLERGVYVLSDEATGKYFTVTKEGNFIPLQLTGKSKVITGFSS 1140
DB 1081 SVKLQWKABYFNGTNVLERGVYVLSDEATGKYFTVTKEGNFIPLQLTGKSKVITGFSS 1140
QY 1141 DGKGIYFCTGCTQAKSAFTVFNNGNTYFEDAGHVMVTNSEYSPNGKDVYRFLPNGIMLSN 1200
DB 1141 DGKGIYFCTGCTQAKSAFTVFNNGNTYFEDAGHVMVTNSEYSPNGKDVYRFLPNGIMLSN 1200
QY 1201 AFYIDANGNTYLYNSKGQMYKGYTKFDVSETDKDGESKVKVFRYFTNEGVMAGVTVI 1260
DB 1201 AFYIDANGNTYLYNSKGQMYKGYTKFDVSETDKDGESKVKVFRYFTNEGVMAGVTVI 1260

QY 1261 DGTQYFGEDEGFOAKDLVTRFKGTTYFDDAHTNGIKDTRWNINGKYYFDDANGVAATGA 1320
DB 1261 DGTQYFGEDEGFOAKDLVTRFKGTTYFDDAHTNGIKDTRWNINGKYYFDDANGVAATGA 1320
QY 1321 QVINGQKLYFNEBDSQVKGGVVKNADGTYSKYKEGFGELVTNNEFFTTCGNVWVYAGANGK 1380
DB 1321 QVINGQKLYFNEBDSQVKGGVVKNADGTYSKYKEGFGELVTNNEFFTTCGNVWVYAGANGK 1380
QY 1381 TVTGAQVINGQHLYPNADGSOVKGVVKNADGTYSKYNASTGERLTNEFFTTCGNNWYYI 1440
DB 1381 TVTGAQVINGQHLYPNADGSOVKGVVKNADGTYSKYNASTGERLTNEFFTTCGNNWYYI 1440
QY 1441 GANGKSVTGEVKIGDDTYFFAKDGKQVKGOTVSAGNGRISYYGDSGKRAVSTWIEIQBP 1500
DB 1441 GANGKSVTGEVKIGDDTYFFAKDGKQVKGOTVSAGNGRISYYGDSGKRAVSTWIEIQBP 1500
QY 1501 VVYVFDKNGLAYPPRVLN 1518
DB 1501 VVYVFDKNGLAYPPRVLN 1518
RESULT 2
Q9XJ5_9STRE PRELIMINARY; PRT; 1512 AA.
ID Q9XJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9XJ5_9STRE PRELIMINARY; PRT; 1512 AA.
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE GTP-S.
GN Name=gTf;
OS Streptococcus criceti.
OG Plasmid pAMI.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]_TaxID=1333;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSSP; P06654; 1MPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9C8C601FC14 CRC64;
Query Match 50.1%; Score 3973.5; DB 2; Length 1512;
Best Local Similarity 52.9%; Pred. No. 2e-170;
Matches 811; Conservative 216; Mismatches 415; Indels 91; Gaps 24;
QY 1 MENKHYKLHKVKKQWVTIAVASVALATVGLSLVTSSTVSADETQDKTIVTOSNCSGTAS 60
DB 1 MERKLRYKLHKVKKQWVTIAVASAGLASVIGA-GAASQTVSADDICNGASASAEQNTSAS 59
QY 61 L---VTSPEATKEADKRTNTKEADY--LTPAKETNAVET--ATTNTQATAEAATTATTA 113
DB 60 QNKEVVDSDAQATDAKATSEQAQVSSVDNTSETDQVQNVGDVGSANNQPEAPAPQAAS 119
QY 114 ---DVAAVFPNKEAVVTTDDAPVTTTEKAEQOPATVKAEEVNVTEVKAPAAALKDSEVEAA 170
DB 120 NNTNTATSEANTNTAVSEAPAAENRTAEK-----GADLSQDEAAAA 162
QY 171 LSLKNIKNIDGKYYVNEBDSGHENFAITVNGQLYF-GKDGALTSSTYSFTPGTTNIV 229
DB 163 LSLDNKIKVDGKYVYVNMADGSYKGFPAITVKGQLLYFDDAETGALSTSTYSFSGQLTPLV 222
QY 230 DGFSINNRAYDSSEASFELIDGLTADSWYRPAIIKDGVTWQASTAEDFRPLLMAWPN 289
DB 233 SDFSINNKAFFUSSAKSFELVDGLTAEWSYRPTKILENGKTWVDSKETDLRPVLTSWMPD 282

Db	439	KNANPDGIRVADVNVNADTLQLTYTNYFNSVYGVNKSEQAALAHISVLEAWSNDNDYNO	498
Qy	521	KTDGAALAMENKORLALLFSLAKPIKERTPAVSPLYNNFTNTORDEKTDWINKGSKAY	580
Db	499	DTNGAALAMNDGLRFLSLYTLRPINERTPGMSLTIKSEYGLTRTK-----	545
Qy	581	NEDGTGVKQSTIGKYNKYGDSNGYVIRAHNNVQDIIAEIIKKEINPKSDGFTITDAE	640
Db	546	-----NDKYGDTQPSYVVRHAHSEVQTVAQIIKEIDFTDGTGFTLDQ	591
Qy	641	MKQAFEYINKDMLSSDKCYTLNIPAAAYAVMLQNMETITRVYGDLYTDDGHMETKSPY	700
Db	592	LKQAFEYINKDMSVKNKHYTHYINPAAYAVMLSNMESVTRVYGDLYTDDGGYMAKSPY	651
Qy	701	YDTIVNLMKSRIKYVSGGQ--AQRSYWLPTDGRQDMSDV-ELYRNEVTSVRYGKDINT	757
Db	652	YDAINTLLRARIRYAAGGQIMEHNSYKPSAAMKAHPDAGNVLGNSEVLVSFRFGQDMS	711
Qy	758	ANDTEGSKYSRTSGQVTLVANNPKMLDOSAKLVNEMGKIHANOKYRALIVGTADGKNF	817
Db	712	ADMTGGKLAKTSGMFTLISNPELELDVNEEIKVNVGKIHAQOYRPLLLTTDDGKLQY	771
Qy	818	TSADAIAAGYVKETDSNGVLTFGANDIKGYETFDMSGFVAVMVPVPGASDNODIRVAPST	877
Db	772	LNDSB---TKLTKIADKDGFIIFKGEIKGYQVEVNGVLSVWVPVGAQADQDIRVAPST	828
Qy	878	EAKKEGELTKATEAVDSOLIVEGFSNFOTIPDGSPPSVYTRKKAENVDLPKSGVGTFS	937
Db	829	AAKGEKAKTYTASQALSESQIIVEGFSNFQDFV--QKDSQYTNKKTAENTDLPKAWGVTFS	886
Qy	938	EMAPQVSADGCTFLDSVLONGYAFADRYDLAMSNNKYGSKEDRLDALKALHKAGIOAI	997
Db	887	EMAPQVSADGCTFLDSIIFENGYAFDRYDLAMSNNKYGSKEDLANALKALHAAGIOAI	946
Qy	998	ADWVPDQIYQLPKCKEVVYTRTDGAKRIADAIIDHSLYVANSKSGKDYQAKYGGEFIA	1057
Db	947	ADWVPDQIYQLPKCKEVVYTRVDNYGRVKVQDPLVEKLYLANTKSGKDFQAKYGGEFIA	1006
Qy	1058	ELKAKYPENPKVMNISTGKPIBDSVKLQWKAEYFNGTNVLRGVGYVLSDBATGKYFTV	1117
Db	1007	ELQKYPENFTTKMISTGKTIDPSVKLEWSAKYFNGTNVLRGTYDILSDEGTYKYFTV	1066
Qy	1118	TKEGNFIPQLTKCKERVITGSSDGKITVFGTSGTQAKSAFVTFNGTYVYFDARGHMT	1177
Db	1067	NEKGDPLPASLTGNKDAKTGFYNDGKGIYITTAGNKARSAFVTEAGNTYIYFDYGHMT	1126
Qy	1178	NSEYSPN--GKDVYRFLPNCIMLSNIFYIDANGTYLYNSKGQYKGYTKFDVSETDKD	1235
Db	1127	----GNVINTKPYIYFLPNCIMLKDAIKODEKRSVYVYKTYGVMYKGRDNEWFAMTDSK	1182
Qy	1236	GKESKVVKPRYFTNEGVMAGVTVIDGFTQYFGEDGFQAKDLVTPK-GKTYIYFDAQTON	1294
Db	1183	GQ----MRFRHFDPRYGFMSGLVTINQNVQYVDENGFOVKGBFVTDQDQTRYFDQSGN	1238
Qy	1295	GIKOTWRNLNGWYIFDANGVAATGAOVINGOKLYEN-EDGSQVKGVVKNADGTYSKYK	1353
Db	1239	LVKQFLNKDGNWYILDDQGLVAKGAQTIKGOKLYFDTKTGVOVKGDFVTDKGNFTFFYS	1298
Qy	1354	EGFGLVTFNEFTTDDGNWYVYAGANGKTVTGAOVINGOHLFYNA-DGSQVKGVVKNADG	1412
Db	1299	GDTGDILILQPFSTGNANFYADENGHVAKGAKTIRGOKLYFDTKTGQAKGFPIDDDKG	1358
Qy	1413	TVSKYNASTGERLTNBFFTT--GDNWYIYGANGKSVTGEVKI-GDPTYFFAKDGQKVGK	1469
Db	1359	V-RYYDADTGTLVNAFLETAKGSNQWYMGADGAVKGNQTIKNQHMVYFDAQGQAKG	1417
Qy	1470	QTVSAGNRLSYIYVYDGSKRAVSTWIEIQPGVYVYFDKNGLA	1511
Db	1418	IIVTDANGRKRYFYDTFTGSRVWNQFVLVN-GNMYFFGYDGA	1458

```
Db 521 SLAVLTK-----QPGQRIDLNLISESVNKERAND-----T 551
Qy 597 KYGDASGNVYFIRAHNNVQDIIAIIKKEINPKSDGFTITDAEMKQAFIYNKMDLSSD 656
Db 552 AYGDITPTYSFVRADHSEVQTVIAKI VKEKIDTNSDGYTFTLDQLDKAFKIYNEDMAKVN 611
Qy 657 KKYTLNIPRAAYVLMOMETITRIVYVYDLYTDDGHYWKTSPPYDIIVNLMKSRIKYVS 716
Db 612 KTYTHYNIPAYALLSNMESVPRVYVYDLYTDDQYMAKSPYDYDAIATMLQGRIAYS 671
Qy 717 GGAQRQSYWLPDTCMDNSDVLYRTNEVTSVRYGKDIIMTANDEGSKYRSYSGQVTLV 776
Db 672 GQSQEEVH-----KVNQNNQILSSVRYGQDILMSADDTQGTDLRSISGLVTLV 718
Qy 777 ANPKMLDOSAKLNVEMGKHANQYKRALIVGTADGINKFTSDADAIAAGYVKETDSNG 836
Db 719 SNDPNLDLGDS-LTVNNGRAHANQAVRPLILGTGDKGVQSVLKDS---TNIVKYTIDANG 774
Qy 837 VLTFGANDIKGYETFDMSGFVAVWVPVPGASDNQDIRVAPSTEAKKEGELTLKATEYDSQ 896
Db 775 NLFTTADDIKGYSTVDMSGVLAVWVPVPGAKDQDVRVAADTNQKADGK-SLKTSAA LDSQ 833
Qy 897 LIYEGFSNFOTIPDGSPPSVYNNRKIAENVDLFKSMGVTSFEMAPQFVSADDTGFLDSVI 956
Db 834 VIYEGFSNFQDF--ANNADYTNKXIAENADFPKLGITSEFEMAPQIVSATDGSFLDSII 891
Qy 957 QNGYAFADRYDLAMSKNNKYGSKEDLRDALKALHAKAGIAIADWPDQIYQLPGEVYTA 1016
Db 892 QNGYAFSDRYDLAMSKNNKYGSKDDLALANLAKALHANGIQAIADWPDQIYQLPGEVYTA 951
Qy 1017 TRTDGAGRKTADAILDHSLYVANSKSGKDYQAKYGGEFLEAELKAKYPENFKNMISTGK 1076
Db 952 KRTNSYGNPFDYAINNALYATNTKSSGSDYQAYGGAFDELKAKYPDMFTVMNISTGK 1011
Qy 1077 PIDDSVKLKQWKAIEYFNGTNVLRGCVYLSLSDATGKYFTVTKEGNFIPLQLTGKKEVIT 1136
Db 1012 PIDPSTKIKQWEAKYFNGTNVLRGAGYVLSLSDATGKYFTVNSGDFPLPASFTGDQNAKT 1071
Qy 1137 GFSSDGKGIYFTGSGTQAKSAFVTFNGNTYTFDARGHMVTNSSEYSPNGKDYRFLPNGI 1196
Db 1072 GFYVDTGTMAYYSTGNKAVNSFIYEGGHYYYFDKGHMVYTSYKAEKGNDYY-FLPNGI 1130
Qy 1197 MLSNAPYIDANGTYLYNSKGQMYKGG--YTKFDVSETDKDGKESKVKVFRYTNBGMVA 1254
Db 1131 QMRDAIYQDAQGSNTYTGRTGILYKGNWYFPFVDPNNANK-----TVFRYFDANNVMA 1183
Qy 1255 KGVTVIDGFTQYFEGDGFQAKDKLVTFKGTYYFDAHTGNGIKDWRNINGKMYFYDANG 1314
Db 1184 IGRNNYIGQYYFDENGFQANGQLLTDGKTHYFDENGAMAKNFVNVGDDWYMDNG 1243
Qy 1315 VAATGAQVINGQKLYFN-EDGSQVKGGVKNADGTYSKYKEGFGELVTNEFFTTDGNVWY 1373
Db 1244 NAVKGQYPVNNQILYFNPETGVQVKQFITDAQRTSYDANDSALKSSGFFTFPNSGDWY 1303
Qy 1374 YA-----GANGKTVTGAQVINGQHLYFNAD-GSQVKGGVKN 1409
Db 1304 YAENGVYVYKGFKOVAENQDQWYFDQTTGKQAKGAARVQGRDLYFNPDSGVQVKGFATD 1363
Qy 1410 ADGTYSKYNASTGERLTNEPFTTGDNNWYIIGANGSKSVTGEVKI----- 1453
Db 1364 ESGNTSFYHGDNGDKVVGFFFTTGNNAWYADNNGNLVKGFBQIDGKWHFDEVTCQOAK 1423
Qy 1454 -----GDDTYFFPAKDGKQVKGQVTSAGNGRISYYYGDSGKRAVSTWIEIQGVVYVFDK 1507
Db 1424 GAALVNGQQLYFVDVDSGIQVKGDFVTDGQNTSYDVNSGDKKXNGFFTTGDNAWYADG 1483
Qy 1508 NG 1509
Db 1484 QG 1485
```

RESULT 5
Q00599_STRSL

```
ID Q00599_STRSL PRELIMINARY; PRT; 1599 AA.
AC Q00599;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-FEB-2005 (T-EMBLrel. 29, Last annotation update)
DE Glucosyltransferase S.
GN Name=gtfk;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=93381463; PubMed=8371114;
RA Giffard P.M., Allen D.M., Milward C.P., Simpson C.L., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and
evolution of the gtf genes of oral streptococci.";
RL J. Gen. Microbiol. 139:1511-1522(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least tow
glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
RL J. Gen. Microbiol. 137:2577-2593(1991).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RA Jacques N.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25975;
RA Giffard P.M., O'Connor D.M., Milward C.P., Simpson C.L., Jacques N.A.;
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and
the evolution of the gtf genes of oral streptococci.";
RL Mol. Microbiol. 0:0-0(1992).
DR EMBL; Z11872; CAA77898.1; -; Genomic DNA.
DR EMBL; Z11873; CAA77901.1; -; Genomic DNA.
DR PIR; S22737; S22737.
DR HSSP; P06653; 1H8G.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Transference.
SQ SEQUENCE 1599 AA; 176480 MW; 24B77869BE152B707 CRC64;

Query Match 47.2%; Score 3743; DB 2; Length 1599;
Best Local Similarity 47.1%; Pred. No. 5.3e-160;
Matches 787; Conservative 234; Mismatches 417; Indels 232; Gaps 26;

Qy 1 MENKHYKLHKVKKQWNTIIVASVALATVGLGLSVTTSSVSADETODKTVTQNSGTTAS 60
Db 1 MENKRYKLHKVKKQWNTIIVASVALATVIGSVATSSLASAETNNS--NGSPSTTTVG 58
Qy 61 LVTSPEATKEADKRTNKTEADVLTPAKETNAVETATTTNTQATAEAATTA-----TTADV 115
Db 59 ENTNPVVEKEVGTTT-----EVANTSNNATTTTTERAEVTDKPKAGTTVQP 101
Qy 116 AVAAPNKEAVTTDAPATTEKAE-----EQPATVKAEEVNVTEVKAP-----EAAKDSSEV 167
Db 102 NSGTTSDRAAAVEVEAKPETTAKPEVATKPEATTATTSSEVAANAGVAAAPTTEKSELSEAEI 161
Qy 168 EAALSINKNI-KNIDGKYVYVNDGSHKENFALTNGQLLYFGKGALTSSTSYSTPPTGT 226
Db 162 KAAVSLDNIKKEKDGKYYLLYLEDGSHKKNFALTNGQVLYFDENGALSSTSYSTYFTQETT 221
Qy 227 NIVDGF SINNRAYDSSEASFELIDGLYLTADSWYRPASIIKDGVTWQASTAEDFRLLIMAW 286
```


CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.

CC -!- SIMILARITY: Contains 6 cell wall binding repeats.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

CC -----

DR EMBL; M29296; AAA36895.1; -; Genomic DNA.

DR EMBL; D88653; BAA36103.1; -; Genomic DNA.

DR EMBL; D88656; BAA36107.1; -; Genomic DNA.

DR EMBL; D88659; BAA36111.1; -; Genomic DNA.

DR EMBL; D88662; BAA36115.1; -; Genomic DNA.

DR EMBL; D89979; BAA36121.1; -; Genomic DNA.

DR EMBL; AB014932; AAN58619.1; -; Genomic DNA.

DR HSSP; P06653; 1GVM.

DR InterPro; IPR002479; Cell wall bd put.

DR InterPro; IPR003318; Glyco_hydro_70.

DR Pfam; PF02324; Glyco_hydro_70; 1.

DR Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal; Transferrase.

KW SIGNAL 1 ? Potential.

FT CHAIN ? Glucosyltransferase-S.

FT REPEAT 1232 1295 1.

FT REPEAT 1296 1359 2.

FT REPEAT 1360 1423 3.

FT REGION 1232 1423 3 X 63 AA approximate tandem repeats.

FT Y -> H (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT I -> V (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT K -> E (in strain MT4467).

FT A -> S (in strain MT4239 and strain MT4245).

FT A -> T (in strain MT4251 and strain MT8148).

FT T -> I (in strain MT4239 and strain MT4245).

FT A -> V (in strain MT4239, strain MT4245 and strain MT8148).

FT A -> V (in strain GS-5 and strain MT4467).

FT A -> V (in strain MT4245).

FT A -> T (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT V -> L (in strain MT4239).

FT D -> N (in strain MT8148).

FT E -> D (in strain MT4239, strain MT4245 and strain MT4251).

FT D -> N (in strain MT4239, strain MT4245 and strain MT4251).

FT Q -> H (in strain MT4245).

FT D -> N (in strain MT4239 and strain MT4251).

FT E -> K (in strain MT4239).

FT F -> P (in strain MT4239).

FT F -> L (in strain MT4239, strain MT4251 and strain MT4467).

FT KKYYTC -> EKEYTL (in strain MT4251).

FT A -> S (in strain MT4239).

FT TQGGSA -> ADKGNDS (in strain MT4251).

FT TDQGS -> ADKGN (in strain MT4239 and strain MT4245).

FT T -> A (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT D -> Y (in strain MT4251).

FT E -> K (in strain MT4245 and strain MT4251).

FT LG -> IR (in strain MT4251).

FT G -> R (in strain MT4245).

FT G -> R (in strain MT4239).

FT H -> Q (in strain GS-5).

FT S -> N (in strain MT4239).

FT Y -> C (in strain MT4251 and strain MT4467).

FT F -> L (in strain MT4467).

FT Q -> P (in strain MT4245).

FT K -> T (in strain MT4245).

FT N -> D (in strain MT4245).

FT G -> D (in strain GS-5 and strain MT4467).

FT R -> R (in strain GS-5).

FT R -> K (in strain MT4467).

FT RYDKNSGNMVKVVTLLANGRRIGIDRWGIARYY -> VY R (in Ref. 1).

FT SQ SEQUENCE 1462 AA; 163388 MW; CE4A279C4D708645 CRC64;

Query Match 44.6%; Score 3532; DB 1; Length 1462;

Best Local Similarity 47.6%; Pred. No. 1.5e-150;

Matches 731; Conservative 253; Mismatches 451; Indels 100; Gaps 31;

QY 1 MENKIHYKLHKVKQWVTIIVASVALATVGLGLSVTTSSVSAD---ETQDKVTVO-SNSG 56

DB 1 METKRYKMYKVKGHVVTIIVAS-GLITL--GTTTIGSSVSAETEQTSDKVVYTKSQSDD 57

QY 57 TTASLVTSPEATKADKATNTKEADVLTPAKETNAVETATTTNTQATAEAAT-ATTADV 115

DB 58 KAASSSQTDAPTKQAQTEQTAQ-----SQANVADTSTSTKTPSPNITTOANSDDK 112

QY 116 AVAAVNPKEAVVTTDAPAVTTEKAEQPAATVAAEVVNTSEVKA-----PEAAKDKSEVEA 169

DB 113 TVTNTKSEEAQTSSE-----RTKQAEQAQATASSQAL-TOAKAELTKQRTAQENKNPVD 167

QY 170 ALSLKNIKNDGKYIYVNEEDGSHKENPAITVNGQLLYFGKO-GALTSSSTYSFTPTTNI 228

DB 168 LAAIPNVKQIDGKYIYIGSDGQPKKNFALTNNKVLFPDKNTGALTDTTSQYQFKQGLTKL 227

QY 229 VDGFSINNRAYDSSSEASPELDGVLTAADSWYEPASIIKDGVTWQASTAEDEFPLLMAMP 288

DB 228 NNDYTPHNQIVNFENTSLETIDNYTADSWYEPKDLKNGKWTWASSSESDRLPLMSWPP 287

QY 289 NYDVTQVNTLYNYSKV-FNLDAKYSTDKQETLKVAAKDIQIKIEQKIQAEKSTQMLRETI 347

DB 288 DKQTOIAYLYNNYQOGLGTGENYTADSSQESLNLAATVQVVKLETIKISQTOQTOWLRDII 347

QY 348 SAFVKTPQWNKETENYKGGGHDHQQGALLYVNDSTRTPWANSYDRLNRTATNQGTI 407

DB 348 NSFVKTPQWNNSQTSSTDSAGEKDHQQGALLYNSDKTAYANSYDRLNRTPTSQTGK- 406

QY 408 DKSILDEQSDPNHMGCFDFLLANDVDLSNPVYQAEQLNQHLYLMNWSGTYMGDKDANFQ 467

DB 407 -----PKYPEDNSSGGYDFLLANDDINSNPVYQAEQLNWLHYLMNYSIVANDPEANFQ 461

QY 468 IRVDAVDNVDAQLQYNTYFREYGYGVNKSSEANALAHISVLEAWSLNDNHYNDKTDGAAL 527

DB 462 VRVDAVDNVNADLLQIASDYLKAHGVDKSEKNAINHLISLEAWSNDNDPOYKNDYKGAQL 521

QY 528 AMENKQRLALLFSLAKPIK-----ERTPAVSPLYNNFTNTORDEKTDWINKDGSKAY 580

DB 522 PIDNKLRLSLVALTRPLEKXDAKNKNEIRSGLEPVITNSLN-----562

QY 581 NEDGTGVKOSTYKYNKYKGDASGNVYFRAHDNNVQDIAEIIKKEINPKSGDFTITDAE 640

DB 563 -----NRSAGEKNSER-----MANYIFIRAHDSVQTVIAKIIKQAINPKTDGLTFLDE 612

QY 641 MKQAFEPKDWLSSDKKYTLNINIPAAVAVMLQNQMETITRVVYVYGDLYTDDGHYMETKSPY 700

```
Db      613  LQOAFKIYNEDRQAKKQTQSNIPAYALMLSNKDSITRLYGDWYSDGQYMAKSPY 672
      :||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy      701  YDTIVNLMSRIKYVSGGOAQRSYWLPDTGDKMNSDVELYRTNEVYTSVRYGKIDMTAND 760
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db      673  YDAIDTLLKARIKYAAGGQDMKITV--SGDKSHMD--WDYTGVLTSVRYGTGANEATD 727
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy      761  TSGSKYSRTSGQVTLVANNPKLNLQSAKLNLEMGIHANQKYRALIVGTAGIKNFTSD 820
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db      728  -QGSEATKTQMAVITSNNPSLKNQDKVIVNMGTAKHNOBYRPLLLTTKDGLSYTS- 785
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy      821  ADAIAAGYKEDTNSCNLFTFGANDIKGYETFDWMSGFVAVVVPVPGASDNODIRVAPSTEAK 880
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db      786  -DAAAKSLRKNDKGELVFDASDIQOYLNPQVSGYLAWVVPVPGASDNODIRVAPSTEAK 844
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy      881  KEGELTLKATEAYDSQLIYEGFSNFQT-IPDGSDPVSVYNNRKIAENVDLFSKSGVTSFPM 939
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db      845  ATGOV-YESSALDSQLIYEGFSNFODFVTKDS--YTNKKIAQNVQLFSKSGVTSFPM 900
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy      940  APQFVSGADGTFLDSVIQNGYAFADRYDLAMSKNNKYGSKEDLRDLKALHKAQIAIAD 999
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db      901  APQVYSEDSGFLDSIIQNGYAFEDRYDLAMSKNNKYGSGQDMINAVKALHKSIGIQTAD 960
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     1000  WPDQIYQLPGKEVWTATTDGAGRKIADAIIDHSLYVANSKSGKDYQAKYGGEPFLAE 1059
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db      961  WPDQIYQLPGKEVWTATRVNDGEYRKDSEIKNTLYAANTKSGKDYQAKYGGEPFLAE 1020
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     1060  KAKYPMFMKVMILSTGKPIDDSVKLQWKAEPFNGTNVLERGVYVLSDEATGKFTVTK 1119
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db     1021  AAKYPSIFNRTQISNGKIDPSEKITAWKAKYFNGTNILGRGVYVLDKNSDKYFELAG 1080
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     1120  EGNFIPLOITGKEKVIITGSSDGKGYTFGTSGTQAKSAFV-TFNGNTYVFAARGHMVTN 1178
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db     1081  NOTYLPKQMTNKE-ASTGFVNDGNGMTFYSTSGYQAKNSFVQDAKGNWYVFDNNGHMVYG 1139
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     1179  SEVSPNGKDYVRLPGLMILSNAPFYIDANGNTYLYNSKQMYKGGYTKEDVSETDKDGE 1238
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db     1140  LQHL-NG-EVQVPLSGVQLRESFLENADSGSKNYFGHLNRSYNGSYFDNDS----- 1190
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     1239  SKVVFERYTNEGVMAKGVTVIDGTFQYFGEQDQFQAKDLVT-FKGKTYVFAHTNGIK 1297
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db     1191  ----KWRYFDASGVMAVGLKTNGNTQYFDQDGYQVKGAWITGSDGKKEYFDDGSGNMAV 1246
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     1298  DTWRN-INGKTYFDANGVAAGAQTNGKLYFNGEDGQVGGVVKVKNADGTYSKYKEGF 1356
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db     1247  NRPANDKNGDWYLYNSDGLALVGQTINGKTYVFGQDGKQIKGKIITD-NGKLKYFLANS 1305
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     1357  GELVTFNEFTDGNVNVYAGANGKTVTGAQVINGQHLVFNADGSOVKGGVKNADGTYSK 1416
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db     1306  GELARNIFATDSQNNWYFSGDGAVTGSGTQTAGKLYFASDGKQVKGSGFV-TYNGKRVH 1364
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     1417  YNASTGERLTNEPFTTGDNNWYIYGANGSKSVTGEVKGIDGDTYFFAKDGKQVKGQTVSAGN 1476
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db     1365  YHADSELQVNRREADKGNWYLYDSNGEALTGSRINGQRVFFTEGKQVKGQDVAYDER 1424
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     1477  GRISYVYDGSGRKAVSTWIEIQGVVYVFDKNGLA 1511
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db     1425  GLLRYVDKNSGNVNVKVVWTLANGRRIGRIDRWGIA 1459
      :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|

RESULT 7
Q54178_STRGN
ID   Q54178_STRGN PRELIMINARY;          PRT; 1577 AA.
AC   Q54178; Q54247;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Glucosyltransferase.
GN   Name=gtfg;
OS   Streptococcus gordonii.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1302;
```

```
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=Challis;
RX      MEDLINE=96157084; PubMed=8586195;
RA      Vickerman M.M., Sulavik M.C., Clewell D.B.;
RT      "Molecular analysis of Streptococcus gordonii glucosyltransferase
      phase variants.";
RL      Dev. Biol. Stand. 85:309-314(1995).
RN      [2]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=Challis;
RA      Minick P., Vickerman M.;
RL      Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; U12643; AAC43483.1; -, Genomic_DNA.
DR      PIR; B41898; B41898.
DR      HSSP; P06653; 1HCK.
DR      GO; GO:0016740; F:transferase activity; IEA.
DR      GO; GO:0009250; P:glucan biosynthesis; IEA.
DR      InterPro; IPR002479; CW_binding.
DR      InterPro; IPR003318; Glyco_hydro_70.
DR      Pfam; PF01473; CW_binding_1; 5.
DR      Pfam; PF02324; Glyco_hydro_70; 1.
KW      Transferase.
SQ      SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 42.6%; Score 3373.5; DB 2; Length 1577;
Best Local Similarity 45.7%; Pred. No. 2.4e-143;
Matches 719; Conservative 228; Mismatches 495; Indels 133; Gaps 28;

Qy      1  MENKHYKHKVKQKWITIAVASVALATVLGSLSVTSSVSAD-----ETQ 46
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db      2  MEKKVHYKHKVKQKWVALAVTSLALVAPKALGLESVIYADDAQNVNKEQSAVQSK 61
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     47  DKTVTQSNST-----TASLVTSF---EATKADKRTNTKEADVLTP 85
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db     62  DSEQITSDKATDSSQLEKVEQASSKETVQASAAITNPANEQTQDKEVETSRDTSRHE 121
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy     86  AKETNAVETATTNTQATAEAAATTATTADVAVAAPVNPKEAVTTTAPAVTTKEAEQPAT 145
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db    122  LTQKTSDDSSSEKSGSQBPVKVADQAESTDKTQALQAKQDSRANDQOEETTENVAK---AT 178
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy    146  VKAEVVNTEVK--APE-AALKDSEVEAALS-----LKNIKNIDGKYVY 186
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db    179  VSDKIATPKKERLPEPQKRESITEKMLAAQQAAPVNTTEHDDVLAHIKTIIDGKKYV 238
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy    187  NEDSGHKNFAITVNGQLLYF-GKDGALTSSTSYSTFTPTTNIVDPGFSINNRAYDSSEAS 245
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db    239  QDGTGVKKNFVELNGKILYFDAETGALVDSNEYQFQQTSSLNNEFTQKNAPYGTGTDKD 298
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy    246  FELIDGYLTADSWYRPPASIIKDGVTWQASTAEDFRPLLMWPNVDTQVNYLYNYSKVPN 305
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db    299  IETVDGYLTADSWYRPPKFIKDGKWTATSTETDLRPLLMWPNDKRTQINLYNMQE-N 357
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy    306  LD-AKYSSTDQKETLKVAAKDIQIKIEOKIQAESTQWLRETISAFVKTQPOWNKETENY 364
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db    358  LGIGAFESKTEQVLLTNVQVQVKRIEIKSEKGTIKWLRTLMSAFVKTPQNNIKTESE 417
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy    365  SKGGGEDHLOGGALLYVNDSTRTPWANSYRRLNRTATNTQGT----IDKSIILDEQSDPNH 420
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db    418  TTGTGNKDLHLOGGALLYTNSDKTSHANSRYLLNRTPTSTQTPKPYFDKS----- 467
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy    421  MGGFDFLLANDVDLSNPVVOAQLNQIHYLMWGSIVMGDKDANFDGIRVDADVNDADM 480
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db    468  NNGYBFLLANDFDNSNPVQVQAEQLNWLHYMNFSGIVANDPTANFDGVDVADVNNADL 527
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy    481  LQLYTNYFREYYGVNKSSEANLAHISVLEAWSLNDHNNDKTDGAALAMENKORLALLFS 540
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db    528  LQIASDYFKSRKYGVSEEEALKHLSILEAWSDNDPDYKDTKGQAQLADNKLRLSLLYS 587
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Qy    541  LAKPIKERTPAVSPLYNNFTNTTORDEKTDWINKDGSKAYNBEDGTVKQSTIGKYNEKYGD 600
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
Db    588  FMRKLSIRS-GVEPTITSLN---DRSTE-----NKNGE 617
      |||:||||:||||:| :||||:| :||||:| :||||:| :||||:| :||||:| :||||:|
```



```
QY 606 VFIADHNNVDIIARIKKKEINPKSDGFTITDAEMKQAFELYNKMDLSSDKKYLNNIP 665
Db 621 IFVRAHSEVQVIADIIKENINPNDGLFTTMDLQKAFKLYNEDMRKADKKYIQFNIP 680
QY 666 AAYAVMLQNMETITRYVYGDLTYDDGHYMETKSPYYDTIVNLMKSRKIKYVSGQQRYSW 725
Db 681 TAAHALMLSKNDSTIRVYVYGDLYTDGQYMEKSPYHDAIDLARARIKYVAGQDMKVY 740
QY 726 LPTDGHKNDSDVELYRTNEVTSYRYGKOIMTANDTEGSKYKRTSGQVTLVANNPKLND 785
Db 741 MGVPREADK-----WSYNGILTSYRYGTGANEATD-EGTAERTQGMVIANNNPKLN 794
QY 786 QSAKLNVEMKIHANOKYFALVGTADGJKNFTSDADAJAAGVVKETSDNGVLTGCANDI 845
Db 795 EWDKLGNNGAHKQNYRPVLLTTKDGISRYLTD-EEVPQSLWKTKDANGILTDFMDNDI 853
QY 846 KGYETFDMSGFVAVVPVUGASNDQDIRVAPSTEAKKEGELTLKATEAYDSQLIYGFNSF 905
Db 854 AGYSNVQVSGYLAVVPVUGAKADQDARTTASKKNASQGV-YESSAALDSQLIYGFNSF 912
QY 906 QIIPDGSPSVYTRNKIAENVDLFSKWGVTSEMAPQFVSADDTGLDLSVIONGVAFADR 965
Db 913 QDPATRDDQ--YTNKVIKKNVLFKEWGVTSFELPPQYVSSQDGTFLDSIIQNGYAFEDR 970
QY 966 YDLAMSKNNKYSKEDRLDALXALHKAGIQATADVDPDQIYOLPGKEVVTATRTDAGRK 1025
Db 971 YDMAMSKNNKYSGLDKLLNALRALHSVNIQALADVDPDQIYNLPKGVEVVTATRVNNGYV 1030
QY 1026 IADAIIDHLSHYVANSKSGDKYQAKYGGFGLAEKAKYPEMPKVMNISITGPKIDDSVKLK 1085
Db 1031 REGAEIKEKLYVANSKTNETDQKYGGAFLDELAKAYEIPERVQISNGQKMTTDEKIT 1090
QY 1086 QWKAETFGNTNVLERGVYVLSDEATGKYFTVTKEGNFIPLOLTGKEKIVITGFSSDGKI 1145
Db 1091 KWSAKYFNGTILGRGAYTVLKDWAENDYLTNRNGEIVLPQLVKNKS-YTGFVSDANGT 1149
QY 1146 TYFGTSGTAKSAFV-TFNGNTYFDARGHMTNSEYSPNGKDVYFELNGMLSNFAYI 1204
Db 1150 KFYSTSGYQAKNSFIODENGNWYFDPKRGYLVGTGA-HEIDGKHVY-FLKNGIQLRDSIRE 1207
QY 1205 DANGNTLYNSKG-QMYKGYTKFDVSETDKDKESKVVKFRYFTNEGVMAGKVTVIDGF 1263
Db 1208 DENGNYQYVDQGAQVLRNYT-----TDGQ-----NRYFDKAGVMARGLVKIGD 1254
QY 1264 TQVFGEDGQAKDLVTFK-GKTYYPDAHTNGI-----KOTW----- 1300
Db 1255 QQFFDENGQVVKGIYSAKDGKLRYPDKDSGNVNRPAQGDNPSPWYFVGVEFAKLITGL 1314
QY 1301 -----RNNGK----- 1315
Db 1315 QKIGQOTLYPDQDGQVKGKIVTLSDKSYRYFDANSSEMAGKPAEGAKNEWYFDKTK 1374
QY 1316 AATGAQVINGQKLYFNEGDSQVKGKGVKNADGTYSKYKEGFGELVNEFTFTDGNWVYA 1375
Db 1375 AVTGLQKIGKQTLFPDQDGQVKGKVVTLADRSIRYFDADSGEMAGKPAEGAKNEWYF 1434
QY 1376 GANGKTVTGAQVINGHLYFNADGSOVKGKGVKNADGTYSKYKNASTGERLTNEFTFTGN 1435
Db 1435 DQTKRAVTGLQKIDKQTLFPDQDGQVKGKIVTLSDKSYRYFDANSSEMAGKPAEGAK 1494
QY 1436 NMYIYGANGKSVTEVKIGDITYFFAKDGKQVKGQTVSAGNRIYYYGDSGKRAVSTWI 1495
Db 1495 EWYIFDQAKAVTGLQVQGOQTLTYFDQDGQVKGKVVDDV-NGVSRVFDANSMDMAKSKI 1553
QY 1496 ETQPGVYVYFDKNG 1509
Db 1554 QLEDGSMWYFDRDG 1567
```

RESULT 9
Q59983 9STRE
ID Q59983 9STRE PRELIMINARY; PRT; 1590 AA.
AC Q59983;

```
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-WAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Glucosyltransferase-I precursor (EC 2.4.1.5).  
GN Name=gtfi;  
OS Streptococcus sobrinus.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI_taxID=1310;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CM2176;  
RX MEDLINE=94146405; PubMed=8312602;  
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;  
RT "DNA sequence of the glucosyltransferase gene of serotype d  
Streptococcus sobrinus";  
RL DNA Seq. 4:19-27(1993).  
RN [2]  
RX PROTEIN SEQUENCE.  
RP MEDLINE=91224988; PubMed=1827439;  
RA Mooser G., Hetta S.A., Faxton R.J., Shively J.E., Lee T.D.;  
RT "Isolation and sequence of an active-site peptide containing a  
catalytic aspartic acid from two Streptococcus sobrinus alpha-  
glucosyltransferases";  
RL J. Biol. Chem. 266:8916-8922(1991).  
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.  
DR PIR; A39841; A39841.  
DR HSP; P06653; 1HCX.  
DR GO; GO:0047849; F:dehydratase activity; IEA.  
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR002479; CW_binding_70.  
DR InterPro; IPR003318; Glyco_hydro_70.  
DR Pfam; PF01473; CW_binding_1; 2.  
DR Pfam; PF02324; Glyco_hydro_70; 1.  
KW Glucosyltransferase; Signal; Transferase.  
FT SIGNAL 1 38  
FT CHAIN 39 1590 glucosyltransferase-I.  
FT SEQUENCE 1590 AA; 175956 MW; C3C83AS7CF3C2B0E CRC64;  
  
Query Match 41.9%; Score 3325.5; DB 2; Length 1590;  
Best Local Similarity 45.0%; Pred. No. 3.5e-141;  
Matches 710; Conservative 239; Mismatches 456; Indels 173; Gaps 33;  
  
QY 1 MENKHYKLHKVKKQWTVIAVASVA-LATVLGGLSVTTSSVSADTQKTVTQNSGTTA 59  
Db 1 MEKNYRFKMKYKKEWVTLVSASATMLASALG-----ASVASAD-----TDTASDD---- 46  
  
QY 60 SLVTSPEATKEADKETNTKEADVLTPAKETNAVETATTTNTQATAEAATATTADVAVA 119  
Db 47 ----SNQAVVTGDTNNQATD-----QTSIAATAT-----SEQSASTDAATQASAA 90  
  
QY 120 VPNKEAVVTTDAPAVTTEKAEQPAVTKAEVNV-----TEVKAPAAALKDSEVEAAL 171  
Db 91 EQTQGTASTDTAAQTNNANE-AKWVPTENENQGTDEMLAEAKNVATAESDSIPSDLA 149  
  
QY 172 SLKNTKNDIGKYYVNEGSHKENPAITVNGQLLTFKGKDALTSSTYSFTPTGTTNVDG 231  
Db 150 KMSNVKQVQDKYVYDQDGNVKNKFNPAVSGDKYIYFDETGYKDTSKVDADSSSAVSQN 209  
  
QY 232 ---FSINNRVYDSSSEAFELIDYLTADSWYRPASIIKDGVTWQASTADPRPLMAWNP 288  
Db 210 ATIFANNRVRVSTSAKNFEAVDNYLTADSWYRPKSIKDGKTWTSGDKDFFPLMAWNP 269  
  
QY 289 NVDTQVNYLNMYSKVFNLDAKYSSTDQKETLKAARADIIQIKIEQIAEKSTQWLRETS 348  
Db 270 DTEKRYNYVMKVVIGIDKTYTETSQADLTAAAEVLQARIEQKITSNNTKWLRATS 329  
  
QY 349 AFVKTPQPNKKTEN-YKGGGEDHLQGGALLVYVDS-RTPWANSDYRLNRATNTQGT 406  
Db 330 AFVKTPQPNWGESEKPY-----DDHLQGALLFDNQDITLPTQSNYRLNRPTNTQGS 384  
  
QY 407 IDKSILDEQSDPNHMGGFDFLLANDVDLSNPVVQAEQLNQIHYLMNMGSIYVMDKDANFD 466
```


[illegible]

Qy	1068	KVMI	STGKEP	IDSVKLQKWKASYFNGTNTVLRGVGVYLSDEATGKYFTVTKEGNFPLQ	1127	
Db	1001	TKQI	STGQAD	IPDSVKIKQWSAKYFNGSNILGRGADYVLSQDASNKLYNVSDDKLFLPKT	1060	
Qy	1128	LTCG	KVITCF	SSDGKGITV-FGTSGTQAKSAFTVTEGNTYTFDARGHVMTNSEYSPNGK	1186	
Db	1061	LLG-Q	VVESGIR	FDGTGVVNSSTTGEKVTDSITEAGNLYYFGQDGYMTGQAQ-NIKGS	1118	
Qy	1187	DVYR	FLPNGL	IMLSNAFYIIDANGNTYLYNSKGYTKGQYTKPDVSETDKDGKESKVKVKFY	1246	
Db	1119	NYV-FL	ANGAALR	NTVYTDAGQGNHYGNDGKRYENGYQQF-----GNDS---WRY	1165	
Qy	1247	FTNE	GWAKGV	TVIDGTFQYFGSDGFOAKDK-LVTFPGKTYTFDAHTGNGIKDTW-RNIN	1304	
Db	1166	FKN-G	YNALGL	LTAVDGHQVYFDKDGVOAKDKIIVITRDGKRYIFDOHNGNAVTTTFVADKT	1224	
Qy	1305	GKWW	YDFANG	VAAATGAQVINGQKLYFNEDSGSVKGVGVKNADGTYSKYKEGFGELVTNEF	1364	
Db	1225	GHYY	LKDGAV	TGAQTVGKQHLYPEANGQQVKQGVFTAKDKGLFYDVDSDGMWNTTF	1284	
Qy	1365	FTTD	GNVWY	YAGANGKTVTGAQVINGOHLVFNADGQVKGGVVKNADGTYSKYNASTGER	1424	
Db	1285	IEDK	AGNWF	YLGKDGAAVTGAQTIKGKLYFKANGQQVKGDIVKDADGKIRYDAQTGEQ	1344	
Qy	1425	LTNE-			-----FFTTGDN	1435
Db	1345	VFNK	SVSVNG	KTYTFYFGSDGTAQTAQNPKGQTFKDGSGVLRYNLEGOYVSGSGMYETAEH	1404	
Qy	1436	NWYI	GANGKS	VTGEVKIGDDTYFFAKDGKQVKGQTVSAGNGRISYYVYDGSGRVAVSTWI	1495	
Db	1405	EWYV	-KSGV	LGTGAQTIGNQRYVFKDNGHQVKGQLVTGNDGKLYRYDANSQGQAFNKS	1463	
Qy	1496	EIQG	VVVYF	DFDKXGLA	1511	
Db	1464	TVNG	KTY-YFG	SDGTA	1478	

RESULT 11

Q55263_9STRE PRELIMINARY; PRT; 1590 AA.

AC Q55263; AC Q5263;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE GiF-I.

GN Name=Glucosyltransferase;

OS Streptococcus sobrinus.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1310;

[1]

RN NUCLEOTIDE SEQUENCE.

RP STRAIN=ATCC 33478;

RC Sato S.;

RA "DNA and amino-acid sequences of water-insoluble-glucan synthetase produced from Streptococcus sobrinus ATCC 33478.";

RT Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).

RL [2]

RN PROTEIN SEQUENCE.

RP MEDLINE=9122498; PubMed=1827439;

RX Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;

RA "Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases.";

RT J. Biol. Chem. 266:8916-8922(1991).

RL EMBL; D63570; BAA09792.1; -; Genomic_DNA.

DR PIR; A39841; A39841.

DR HSP; P06653; IGWM.

DR GO; GO:0009250; P;glucan biosynthesis; IEA.

DR InterPro; IPR002479; CW binding.

DR InterPro; IPR003318; Glyco_hydro_70.

DR Pfam; PF01473; CW binding 1; 3.

874 TDGQFLDSVIQNGYAFTRDYLGMKANKYGTADQLVKAIKALHAKGLKVMADWVPDQMY 933

1007 QLPGEKVVATRTDAGRKIADAIIDHSLVYVANSKSGKDYOQAKYGGEFLAELKAKYPYM 1066

934 TFPQEVVTVTRDKEGKPIAGSINHSLVYVDTKSGEDDYQAKYGGAFIDELKEKYPEL 993

1067 FKNVMTSTGKPIDDSVKLQWKABEYFNGTNVLRGCVGYVLSDBATGYFTVTKEGNFPL 1126

994 FTKQISTGQAIIDPSVKIKQWSAKYFNGSNILGRGADYVLSDDQVSNKYFNVA5DTLPLPS 1053

1127 QLTGKEKVIKTPSSDGKGLTY-FTSGTQAKSAFVTFNGNTYYTFDARGHNVNTSEYSPNG 1185

1054 SLUGK-VWSGIRYDGKGYIYNSSATGDQKASFITEAGNLVYFGDKGYMVTGAQ-TING 1111

1186 KDVIYRFLPNCIMLSNFIYDANGNTLYLNSKGMQYKGGYTKFDVSETDKDGKESKVVKFR 1245

1112 AN-YFLENGTALRNTIYTDACQNSHYANDGKRYENGQQF-----GND-----WR 1157

1246 YFTNEGMAKGVTVIDGFTQYFGBDGFQAKDK-LVTFKGTYYTFDAHTNGIKDWTW-RNI 1303

1158 YF-KDGNMAVGLTVDGNTQYFEDKCGVQAKDKIIVTRDQGVRYFDQHNGNAVTNTFIADK 1216

1304 NGKWWYFDANGVAATCAOVINGOKLYFNEBDSGVKGVVKNADGTSYKYEKGEGELVTNE 1363

1217 TGHWYLGKDGVAVTGAQTVGKQKLYFEANGEGOVKGFVTSHEGKLYFYDVDSDGMWTD 1276

1364 FFTTDCNVVYAGANKTVTGAOVINGOHLNFNADGSQVKGGVVKNADGTSYKYNASTGE 1423

1277 FIEDKAGNWFYLGKDGAAVSGAQTIRGKLYFKAYGQGVKQGVKGDIVKGTGKIRYDAKSGE 1336

1424 RLUNE-----PFTTG 1433

1337 QVENKTVKAADGKTYVIGNGVAVDPSVVKGTQFKDASGALRFYNLKGQLVGTSGWVETA 1396

1434 DNNWYVIGANGSVTGEVKGIDDTYFPKADQGVKQKQTVSAGNGRISYYVYDGSKGKAVST 1493

1397 NHDWVYI-OSGKALTGEOQTINGOHLFYFKEDGHQVKGQLVTRTDGKRYRYDANSGDQAFNK 1455

1494 WIEIQGVVYVYFDKNGLA 1511

1456 SVTVNGKTY-YFGNDGTA 1472

RESULT 12

GF2_STRDO STANDARD; PRT; 1592 AA.

AC P27470;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 13-SEP-2005 (Rel. 48, Last annotator update)

DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranase)

DE (Sucrose 6-glucosyltransferase).

OS Streptococcus downei (Streptococcus sobrinus).

OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1317;

[1]_

RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].

RC STRAIN=6715 / Serotype G;

RX MEDLINES=91123227; PubMed=1704006;

RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,

RA Kagawa H.;

RT "peptide sequences for sucrose splitting and glucan binding within

RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan

RT synthetase).";

RL J. Bacteriol. 173:989-996(1991).

CC -!- FUNCTION: Production of extracellular glucans, that are thought to

CC play a key role in the development of the dental plaque because of

CC their ability to adhere to smooth surfaces and mediate the

CC aggregation of bacterial cells and food debris.

CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-

CC fructose + (1,6-alpha-D-glucosyl) (n+1).

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
 CC forms of glucans.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -!- SIMILARITY: Contains 16 cell wall binding repeats.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL; D90213; BAAL4241.1; -, Genomic DNA.
 DR HSSP; P06653; IGVM.
 DR InterPro; IPR002479; Cell wall bd put.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW binding_1; 3.
 DR Pfam; PF02124; Glyco_hydro_70; 1.
 KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
 FT SIGNAL 1 38 Potential.
 FT CHAIN 39 1592 Glucosyltransferase-1.
 FT REPEAT 1093 1142 1.
 FT REPEAT 1158 1207 2.
 FT REPEAT 1222 1272 3.
 FT REPEAT 1287 1337 4.
 FT REPEAT 1402 1451 5.
 FT REPEAT 1514 1563 6.
 FT REPEAT 1577 1592 7 (incomplete).
 FT REGION 39 1044 Catalytic (approximate).
 FT REGION 1093 1592 7 X tandem repeats.
 FT REGION 1093 1592 Glucan-binding (approximate).
 SQ SEQUENCE 1592 AA; 176168 MW; BC0A66D079351ECF CRC64;
 Query Match 41.6%; Score 3296.5; DB 1; Length 1592;
 Best Local Similarity 44.4%; Pred. No. 7.1e-140;
 Matches 702; Conservative 240; Mismatches 463; Indels 175; Gaps 32;
 QY 1 MENKIYKHLKVKQWTVIAVASA-LATVLGGLSVTTSSVSADETDKTVTQSNSGTTA 59
 DB 1 MEKNVRFKMKVKRWVTLVSASATMLASALGASVASADTDATSDSDSNQTVV----- 52
 QY 60 SLVTSPEATKADKRNTEADVLTPAKETNAVERATNTTQATAEAATATTADAVAA 119
 DB -----TGDTTNNQATD-----QTSIAATAT-----SEQASTDAATDQASAA 90
 QY 120 VFNKEAVVTTDPAVTTKAEQATVKAENVN-----TEVKAPEAALKDSEVEAAL 171
 DB 91 EQTQGTASTDTAAQTITTWANE-AKWVPTENENQGTDEMLAEAKVATAESDSIPSDLA 149
 QY 172 SLKNTIKRIDKYYVNEGSHKENFAITVNGQLLYFGKDGALTSSTYSFTPTGTTNIVDG 231
 DB 150 KMSNVKQVQKYYVVDQGNVKNFPAVSVDKIIYFDETGAVKOTSVDADKSSSAVSQN 209
 QY 232 ---FSINNRAYSSEASPELIDGYLTADSWYRPASIIKDGVTWQASTAEDFRPLMAWNP 288
 DB 210 ATIFAAANRAYSTSAKNFAVNYLTADSWYRPKSLIKDKGTWETSGKDDFRPLMAWNP 269
 QY 289 NVDVTQVNYLNYMSKVFNLDAKYSSTDKQETLKVAAKDIQIKIEQIOAKSKTOWLRETIS 348
 DB 270 DTETKENYVYNNKVVIGIDKTYTAETSQADLTAAAEVQARIEQKITSNNYKWLREALS 329
 QY 349 AFVKTPQPNKETEN-YSKGGGEDHLOGGALLYVND-SRTPWANSYRRLNRTATNQCT 406
 DB 330 AFVKTPQPNKGESEKPY-----DDHLQNGALLFDNQDTLPTDQSNRYLRLNRTPTNQGS 384
 QY 407 IDKSLIDGSDPNHMGCFPELLANDVDLSNPVQAEQLNQIHLYLMMWGSIVMGDKDANFD 466
 DB 385 LDSRFTYFNDDP--LGGYDFLLANDVDNSNPVQAEQLNWLHYLLNFGSIYANDADANFD 442
 QY 467 GTRVADVNDVADMLQLYNTYREYGVNKSSEANALAHISVLEAWSLNDNHYNDKTDGAA 526
 DB 443 SIRVDAEDNVADQLQISDYLKAAYGIDKNNKANNHVSIVAEWSNDNTPYLHDDGNL 502

QY 527 LAMENKQRLALLFSLAKPIKERTPAVSPLYNNTFNTTQORDEKTDWINKDGSKAYNEDGTV 586
 DB 503 MNMDKFRFLSLMLWSLAKPTDVR-S-GLNPLIHNSLVDREVDDR-----EVEVTV 548
 QY 587 KQSTIGKYNKYGDASGNVVFIRAHNNVQDIIASIIKKEINPKSGDGTITDAEMKQAPE 646
 DB 549 P-----SYSPARAHDSQVQDIIRDIIKAEINPNSFGVSTQEEIDQAPK 592
 QY 647 IYNKMLSSDKKYTLNNIPAAAYAVMLQNMETITRYVYGLYTDGHHYMETKSPYDTIYN 706
 DB 593 IYNEDLKSKDKKYTHYNVPLSYTLLLTNKGSIPIRVYVYGDMTDDQGMANKTVWYDAIS 652
 QY 707 LMSRIKVVSGQAOORSYWLPTDQKMDNSDVELYTNVYTSVRGKDMINTANDTEGSKY 766
 DB 653 LLKARMKYVAGQAMQNYQI-----NGBEILTSVRYGKALKQSD-KGDAT 697
 QY 767 SRTSGOVLTVANNPKLNDQSAKLVEMGKIHANOKYRALIVGTADGINKFTSDADAJAA 826
 DB 698 IRTSGVGVGMQPNFSLDGKV-VALNMGAAHANQERYALMWSTKDGVAITYATDADASKA 756
 QY 827 GYVKETDSNGVLTFGANDIKGYETPDMSGFVAVVVPVGASDNQDIRVAPSTEAKKEGELT 886
 DB 757 GLVKRTDENGILYFLNDDLKGVANPQVSGFLQVWVPVGAADQDIRVAASDTASTDGK-S 815
 QY 887 LKATEAYDSOLIIYEGFSNFOITPDGSDPSVYTNRKIAENVDLPKSWGVTSPEMAFQFVSA 946
 DB 816 LHQDAAMDSRYMFEFGFSNFSQF--ATKEEYEVNVVIANNVDFKFSWGIITDFEMAPQYVSS 873
 QY 947 DDGTFLDSVQNGYAFADRYDLAMSKNNKYGSKEDLRDALKALHAKAGIAIAIDWPVQIY 1006
 DB 874 TDGQVQDSVIQNGYAFTRYDRLGMSKANKYGTADQLVRAIKALHAKGLKVMADWPDQMY 933
 QY 1007 QLPQKEVVTATRTDAGRKIADAIIDHSLYVANSKSSGKDYQAKYCGEEFLABELKAKYPEM 1066
 DB 934 TFPQEVVTVTRTDKFGKPIAGSQINHSLYVTDTKSSGDDYQAKYGGAFDELDEKYPPEL 993
 QY 1067 PKVNMISTGKPIDDSVKLQWKAEPYFNGTNVLERGVYVLSDEATGKYFTVTKEGNFIP 1126
 DB 994 FTKQWSTQQAIDPSVKIKOMSAKYFNGSNILGRADYVLSQVSNKYNFVASDITLFLPS 1053
 QY 1127 QLTGKEKVTGSSDGKIGITY-FGTSGTQAKAFVTFNGNTYTYFDARGHMVTNSEYSPNG 1185
 DB 1054 SLLGK-VESGLRYDGKGYIYNSSATQDVKAASFITEAGNLVYFGKDGVMVTAQ-TING 1111
 QY 1186 KDVRFLPNGLMISNAFYIDANGNTYLYNSKQOMY--KGGYTKFDPVSETDKDGKESKVVYK 1243
 DB 1112 AN-YFFLENGTALRNTIVTDAQNSHYVANDCKRYENENGYQQF-----GND----- 1157
 QY 1244 FRYFTNEGVMAGVTVVTDGFTQYFGEDGFOAKDK-LVTFPGKTYYPDAHTGNGIKDTW-R 1301
 DB 1158 WRYF-KDGNMAYGLTVDGNVQYFDKQGVQAKDKIIVTRDGRVRYFDQHNAGNAVNTFTA 1216
 QY 1302 NINGKMYFDANGVAATGAQVINGOKLYFNEBDSQVKGGVKNADQTYSKYKEGFGELVT 1361
 DB 1217 DKTGHWYILGKDGVAVTGAQTVGKOKLYFEANGQQVKGDFVTSDEGLYFYDVDSDMW 1276
 QY 1362 NEFFTDTGNNVYIYAGANGKTVTGAQVINGOHLVFNADGSQVKGGVKNADGTYSKYNAST 1421
 DB 1277 DTFIEDKAGNWPFLKDGAAVTGAQTIHQKLYFRANGQQVKGDIVKGTGDKIRYDAKS 1336
 QY 1422 GERLTNE-----
 DB 1337 GEQVFNKTYKAADGKTYVIGNDGVAVDPFSVVKQGTFKDASGALRYNLKQLVTGSGWYE 1396
 QY 1432 TGDNNWYIYGANGKSVTGEVKIGDDTYRPAKDGKQVKGQTVSAGNGRISYYVYDGSKRAV 1491
 DB 1397 TANHDMVYI-QSGKALTGEQTINGOHLIFPKDGHQVKGQLVGTGDKVRYYYDANSDDQAF 1455
 QY 1492 STWIEIQGVYVYFDKNGLA 1511
 DB 1456 NKSVTVNGKTY-YFGNDGTA 1474

RESULT 13
 GTFB STRMU
 ID GTFB STRMU STANDARD; PRT; 1476 AA.
 AC P08987; O69381; O69384; O69387; O69390; O69396;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 GN Name=gtfB; OrderedLocusNames=SMU.1004;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TaxID=1309;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=GS-5;
 RC MEDLINE=87308013; PubMed=3040685;
 RA Shiroza T., Ueda S., Kuramitsu H.K.,
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
 RL J. Bacteriol. 169:4263-4270(1987).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
 RC MT4467 / Serotype e, and MT8148 / Serotype c;
 RX MEDLINE=98231643; PubMed=9570124;
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
 RA Kimura S., Hamada S.,
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS Microbiol. Lett. 161:331-336(1998).
 RN [3]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=UA159 / ATCC 700610 / Serotype c;
 RX MEDLINE=22255063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
 RT pathogen."; Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
 CC -1- FUNCTION: Production of extracellular glucans, that are thought to
 CC play a key role in the development of the dental plaque because of
 CC their ability to adhere to smooth surfaces and mediate the
 CC aggregation of bacterial cells and food debris.
 CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
 CC fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
 CC forms of glucans.
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -1- SIMILARITY: Contains 10 cell wall binding repeats.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL; M17361; AAA88588.1; -; Genomic DNA.
 CC EMBL; D88651; BAA26101.1; -; Genomic DNA.
 CC EMBL; D88654; BAA26105.1; -; Genomic DNA.
 CC EMBL; D88657; BAA26109.1; -; Genomic DNA.
 CC EMBL; D88660; BAA26113.1; -; Genomic DNA.
 CC EMBL; D89977; BAA26119.1; -; Genomic DNA.
 CC EMBL; A8014940; AAN58705.1; -; Genomic DNA.
 CC FIR; B33135; B33135.
 CC HSSP; P06653; 1H8G.
 CC InterPro; IPR002479; Cell_wall_bd_put.

DR InterPro; IPR003318; Glyco_hydro_70.
 DR PFam; PF01473; CW_binding_1; 4.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
 KW Transferrase.
 FT SIGNAL 1 34 Potential.
 FT CHAIN 35 1476 Glucosyltransferase-I.
 FT REPEAT 1097 1130 A repeat.
 FT REPEAT 1161 1210 1.
 FT REPEAT 1225 1275 2.
 FT REPEAT 1290 1340 3.
 FT REPEAT 1355 1405 4.
 FT REPEAT 1420 1470 5.
 FT REGION 35 1051 Catalytic (approximate).
 FT REGION 1097 1476 Glucan-binding (approximate).
 FT REGION 1161 1470 S X tandem repeats.
 FT VARIANT 62 62 S -> T (in strain MT4239).
 FT VARIANT 65 65 T -> I (in strain GS-5).
 FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245,
 FT strain MT4251, strain MT4467 and strain
 FT MT8148).
 FT VARIANT 78 78 Q -> P (in strain MT4251).
 FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245,
 FT strain MT4251, strain MT4467 and strain
 FT MT8148).
 FT VARIANT 89 89 S -> F (in strain MT4251).
 FT VARIANT 168 168 K -> N (in strain MT4251).
 FT VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and
 FT strain MT8148).
 FT VARIANT 399 399 N -> R (in strain MT4239).
 FT VARIANT 474 474 I -> T (in strain MT4239).
 FT VARIANT 512 512 K -> R (in strain MT8148).
 FT VARIANT 519 519 F -> Y (in strain MT8148).
 FT VARIANT 701 701 T -> I (in strain MT8148).
 FT VARIANT 708 708 A -> V (in strain MT8148).
 FT VARIANT 938 938 F -> L (in strain MT8148).
 FT VARIANT 952 957 FGKPE -> YGTPVA (in strain GS-5, strain
 FT MT4239 and strain MT4467).
 FT VARIANT 963 964 SV -> NT (in strain GS-5, strain MT4239
 FT and strain MT4467).
 FT VARIANT 968 970 ADS -> VDG (in strain GS-5, strain MT4239
 FT and strain MT4467).
 FT VARIANT 1086 1086 A -> T (in strain MT4239).
 FT VARIANT 1158 1158 S -> N (in strain MT4239).
 FT VARIANT 1163 1163 H -> Y (in strain MT4251).
 FT VARIANT 1168 1168 E -> K (in strain MT8148).
 FT VARIANT 1182 1182 Y -> C (in strain MT8148).
 FT VARIANT 1234 1234 A -> P (in strain MT4239).
 FT VARIANT 1263 1263 R -> H (in strain GS-5 and strain
 FT MT4467).
 FT VARIANT 1263 1263 R -> P (in strain MT8148).
 FT VARIANT 1264 1264 Y -> H (in strain GS-5, strain MT4239,
 FT strain MT4467 and strain MT8148).
 FT VARIANT 1272 1272 S -> G (in strain GS-5, strain MT4239,
 FT strain MT4467 and strain MT8148).
 FT VARIANT 1329 1329 H -> Y (in strain GS-5 and strain
 FT MT4467).
 FT VARIANT 1394 1394 Y -> H (in strain GS-5, strain MT4239,
 FT strain MT4467 and strain MT8148).
 FT VARIANT 1402 1402 S -> G (in strain GS-5, strain MT4239,
 FT strain MT4467 and strain MT8148).
 FT VARIANT 1459 1459 Y -> H (in strain MT4467).
 FT VARIANT 570 570 R -> A (in Ref. 1).
 FT CONFLICT 800 817 ADQDVRVAASTAPSTDGK -> LIKMFALRLARPHQOMA
 FT (in Ref. 1).
 FT CONFLICT 1310 1310 H -> L (in Ref. 1).
 SQ SEQUENCE 1476 AA; 165847 MW; 9C6E09F731B4CBFC CRC64;

Query Match 40.9%; Score 3239; DB 1; Length 1476;
 Best Local Similarity 45.5%; Pred. No. 2.5e-137;
 Matches 700; Conservative 228; Mismatches 474; Indels 136; Gaps 30;
 QY 1 MENKHYKLHKVKQWTVIAVAS--VALATVLGGLSVTTSSVSADFTQDKTVTQNSGTT 58

Db		1	MDKVRYLKRVKRWTVSVASAVMTLTLISGGL-----VKADSNESKSOISNDSENTS	54
Qy		59	ASLVTSPÉATKEADKRNTEADVLTPAKETNAVETATTTNTQOATAEAATATTATTADVAVA	118
Db		55	-----VVTANESNVTTVTSKQEAASSQTNHTVTTISSSTS	91
Qy		119	AVPNKEAV---VTTDAPAVTTEKAEQOPATV---KAEVNVNTEKAEPA---ALKDS---	165
Db		92	VWNPKEVVSNPYTVGETASNGEKLQNTTTVDKTSAAANNISKQTTÉADTDVIDDSNAA	151
Qy		166	EVEAALSNIKNIDGKYVYNVEDGSHENFAITVNGQLLYFGKDGALTSSSYSYFTPG-	224
Db		152	NLQILEKLPNVKEATDGYKYVYDNNKGVRTFTLTJADGKILHFDGTGAYTDTSDTVNKOI	211
Qy		225	TTNIVDGSFINNRAYDSSEASFELIDGVLADSNYRPASIIKOGVTWQOASTAEDFRPLM	284
Db		212	VTRTSNDYKKYNQVYDYSAGSFEHVDHYLTAESWYRPXYILKOGKWTQSTEKDFRPLM	271
Qy		285	AWPNVDTOVNYLNYMSKVFNLDAKYSTDKQETLKVAAKDIQIKIBEQIKAEKSTOWLR	344
Db		272	TWMPSEQTORQYVNYMNAQLGINKYTDTSNQLQNLIAAATIQAKIEAKITLTKNTDWLR	331
Qy		345	ETISAPVKTOPQNNKETNTSKGGEDHLOGGALLYVNDSR-TPWANSDYRRLNRTATNQ	403
Db		332	QTISAFVKTOSAMNSDSEK---PFDDHLONGAVLYDNEGKLTTPYANSNRYILNRTPTNQ	387
Qy		404	TGTIDKSILBQSDP-----NHMGGPFLLANDVDSLNPVQVABQLNQHLYLNMWGSIVM	458
Db		388	TG-----KKDPRYTAONTIGGYEFLLANDVNSNPVQVABQLNWLHFLMNFIGNIYA	438
Qy		459	GDXDANPDGIRVDVADNVADMLQLYNTNYFREYYGVNKSANALAHITSVLEAWSLNDHY	518
Db		439	NDPANFDSIRVDVADNVADLQIAGDYLKAAKGIIHQNDKQANDHLSILEANSNDNTPY	498
Qy		519	NKDTDGAALAWENKQRIALLFSLAKPIKERTPAVSPLYNNTF-NTTORDEKTDWINKDGS	577
Db		499	LHDDGDNMINMDNKLRLSLFLSLAKPLNQRS-GMNPILTNSLVNRTDDNAET-----	549
Qy		578	KAYNEDGTVKQSTIGKNEYKVGDSAGNYVFIIRAHDNNVQDIIAEIIKKEINPKSDGFTIT	637
Db		550	-----AAPVSPFIRAHDSEVQDLIRDIIKAEINPNVGVSYFT	587
Qy		638	DAEMKQAFEYINKDMLSSDKKYTLNNITPAAYAVMLQNNETITRVYGDLYTDGGHYMETK	697
Db		588	MEELKKAFEIYNKDLLATEKYTHYNTALSVALLLTNKSSVPRVYVYGDMTDDGOYWAHK	647
Qy		698	SPYYDTIVNLMKRIKYVSGQAOQRSYWLPTDGHKMDNSDELVRTNREVVYTSVRYGKDIMT	757
Db		648	TINYEAIETLLKARIKYVSGQAMRNQ-----QVGNS-----EIIITSVRYGKGALK	693
Qy		758	ANDTEGSKYRSTSGOVLTVANNPKLINDQSACLNVENGKIHANQYIPALIVGTADGJKNF	817
Db		694	ATDT-GRDTRTSGVAVIEGNNPSRLKASDRVVVVMNNGAAHQNQAYRPLLLTTDNGIKAY	752
Qy		818	TSADADAAGVYKFTDNGVLTFGCANDIKGYETPDMSGFVAVVPVPGASDNQDIRVAPST	877
Db		753	HSQOE--AAGLVRYTNRGELIFTTADIKGVANPQVSGYLGWVPVPGAAADQDVVVAAST	810
Qy		878	EAKKEGELTLKATEBAYDSQLIYEGFSNFQITPDGSDPSVYTNRKIAENVDLFSKSWGYSF	937
Db		811	APSTDDK-SVHONAAALDSRVMEFGFSNFQAF--ATKKEEYTNVVIKXNVKFAEWGVTDF	867
Qy		938	EMAPQFVSADGTFLDSVIONGYAFADRYDLAMSKNNKYKSKEDRLRALKALHWAGIQAI	997
Db		868	EMAPQVYSSTDGSFLDSVIONGYAFTDRYDLIGISKPNKYGTADDLVKAIIKALHSKGIKVM	927
Qy		998	ADWVPDQIYOLPGKEVVYATRTDAGACKIADAIIDHSLXYVANSKSGDKQOAKYGGFELA	1057
Db		928	ADWVPDQWYAFPEKEVVYATRVDFKGPVSGQIKSVLYVADSKSGDKQOAKYGGGAFLE	987
Qy		1058	ELKAKYPEMFKVNNIISTGKPIDDSVKLQKOWKAEYFNGTNTVLERGCVYLVSDÉATGKYFTV	1117

988 ELQAKYPELPARKQIISTGVPMDSVKIKOWSAKYFNGTNI LGRGAGVVLKQDQATNTVFNI 104

1118 T--KEGNFIPLOQTGKEKIVITGESSDGKITGYFTGTSQOAKSAFVTFNGNTYYFDARGHM 1175

1048 SDNKEINFLPKTLLNQDSQV-GFSYDGKGYYVYSTGYQAKNTFISBGDKWYFDDNNGYM 1106

1176 VTNSEYSPNGKQVYRFLPNGIMLSNAFYIDANGNTYLYNSKGOMYKGYTKFDFVSETDKD 1235

1107 VTGAQ-SINGVNYV-FLSNGLOLRDAILKNEEDGTAYYGNDRRYENGYQF----- 1156

1236 GKESKVVKRYFTNEGVMAGVTVIDGFTQYFCGEDGFOAKDKLV-TFKGTYFYFDAHTGN 1294

1157 --MSGV--WRHF--NNGEMSVGLTVIDGQVQYFDPMGVOAKGKFTVTTADGKIRVFDKOSGN 1211

1295 GIKDWT-RNINQKWIYFDANGVAATGAQVINGOKLYFNEDGSOVKGGVVKNNADGTSYKYK 1353

1212 MYNRRIENEEGKWL YLGEDAAVTGSGTINQHLHYFRANGVQVKGEFVTDYGRISYVD 1271

1354 EGFGEVLTVNEFFTDDGNVWYAGANGKTVTGAQVINGOHLHYFNADGSOVKGGVVKNNADGT 1413

1272 SNSGDQIRNRFVRNAQGFYFDNNGYAVTGARTINGOHLHYFRANGVQVKGEFVTDHRGR 1331

1414 YSKYNASTGERLTNEFFTDDNNWYIYGANGSKVTGEVKIGDDTYFEAKDGKOVKGQTVS 1473

1332 ISYDNGSGDQIRNRFVRNAQGFYFDNNGYAVTGARTINGOHLHYFRANGVQVKGEFV 1391

1474 AGNGRISYYDGSCKKXAVSTWIEIQPVVYVDFKNGLA 1511

1392 DRYGRISYDNSGDDQIRNRFVRNAQGFYFDNNGYA 1429

RESULT 14

GTFC_STRMU STANDARD; PRT: 1455 AA.

AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;

DT 01-NOV-1988 (Rel. 09, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Glucosyltransferase-SI precursor (BC 2.4.1.5) (GTF-SI)

DE (Dextranucrase) (Sucrose 6-glucosyltransferase).

GN Name:gtfc; Ordered locus Names=SMU.1005;

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

RP [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=GS-5;

RC MEDLINE=891137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;

RX Ueda S., Shiroza T., Kuramitsu H.K.;

RT "Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.";

RL Gene 69:101-109(1988).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RP STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,

RC MT4467 / Serotype e, and MT4148 / Serotype c;

RC MEDLINE=98231643; PubMed=9570124;

RX Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;

RA "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";

RL PEMS Microbiol. Lett. 161:331-336(1998).

RN [3]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=UA159 / ATCC 700610 / Serotype c;

RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;

RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic C., Chang J.,

RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,

RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,

RA Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental

RT pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

Qy	502	LAHLSVLEAWSLNDHNHNDKTDGAAALAMENKQRLALLFSLAKPIKERTPAVSPLYNNFTN	561
Db	589	IAHLSLSLEAWSYNDHQNKDTGQAQLSDINPLRETLTTLTFLRKSNYR-GSLERTITNSLN	647
Qy	562	TTQDEKTDWINKDGSKAYNEDGTVKQSTICKYNEKYGDASGNTVVFIRAHNDNNQDIIAE	621
Db	648	NRSSE-----QKTPRDANVIFRAHDSEVOAVLAN	678
Qy	622	IIKKEINPKSDGFTITDAEMKQAFEIYNNKOMLSSDKKYTLNNIPAAVAVMLQNMETTIV	681
Db	679	IIKQINPKTDGFTTMDLQKQAFEIYNADIADAKKKTQYQNI PAAVATMLTNKDSITRV	738
Qy	682	YYGDLYTDDGHYMETKSPYDTIYNLMKSRIKYUSGGQAQBSYWLPTDGGKWDNSDVELYR	741
Db	739	YYGDLFTDDGGYMAEKSPYNAIDALLARIKYVAGQ-----DMVKY	782
Qy	742	TN--EVYTSVRYGKDIIMTANDEGSKYSRTSGQVTLVANPNKMLDOSAKLNVEMGKIHA	799
Db	783	LNGYEIMSSVRYGKAEENQL-CTAETRNQGMVLVTANRPMKLGANDRLVNMGAHAK	841
Qy	800	NQKTRALI VGTADGIGNFTSDADAIAGYVKETDSNGVLTFGANDIKGYEFTFDMSGFVAV	859
Db	842	NQAYRPLLKSKTGLATYLDSD--VPAGLVRYTDQGNLTFTADDIAGHSTVEVSGYLAV	900
Qy	860	WVPV GASDNQDIRVAPSTEAKKEGELTKATEAVDSOLIVYEGESNFQTI PDGSDPSVVTN	919
Db	901	WVPV GASENQDARTKAS--STKGEQVFESAAALDSQVIYEGESNFQDFV--KTPPSQYTN	956
Qy	920	RKIAENYDLFKSMGVTSEFEMAPQVPSADDTGLDLSVIQNGYAFADRVDLAMSNNKYGSK	979
Db	957	RVIAQNAKLFKEWGIITSEFAPQVYSSQDGTFLDSIIENGYAPEDRYDIAMSKNNKYGSL	1016
Qy	980	EDLRDALKALKHAGIQAIADWPDOIYQLPGKEVVTATRTDGGAKRTIADAIIDHSLYVAN	1039
Db	1017	KDLMADALRALHAEGISAIADWPDOIYNLPGEVVTASRTNSYSGTPRPNAEINSLYAAK	1076
Qy	1040	SKSGKDQYQAKYGGEFELAEKAKYPKPMKVNMLSTGKPI DDSVKLKQWKAEEYFNGTNVLE	1099
Db	1077	TRTFGNDFOQKYGGAFLDELKAKYPAIFERVOYISNGRKLTTNEKTIQWSAKYFNGSNIQ	1136
Qy	1100	RGVGYVLSDBATGYFTVTKEGNFIPQLQLTGKEKVIYTFPSDGGKGIYTFGTSGTOAKSAF	1159
Db	1137	TGARYVLQDNATNQYFSKAGQTFLPKQMT--EITGSGPRVRGDDQYLSIGGYLAKNTF	1194
Qy	1160	VTENGNT-YVFDARGHMVTHNSEVSPNGKDVYRFLPNCIMLSNAPFYDANGWTVLYNSKG-	1217
Db	1195	IQVGANQWYFYFDKXNGNMVTEGBQVJDGKK--YFFLDNLGLQLRHVLROGSDGHVYYPDPKV	1252
Qy	1218	QMYKGGYTKPDVSBETDQDKGESKVKVPFRYPTNEGVMAGVTVJDGFTQYFGE-DGFOAKD	1276
Db	1253	QAFNGFY-----DFAGPRQDV--RYFDGNGQWYRGLHDMYGTTFYFDEKTIQAKD	1301
Qy	1277	KLVTFP-KGKTYYPDAHTGN-GIKDTHWNIGK-WYYPDANGVAATGAQVINGOKLYNFED	1333
Db	1302	KFIRFADGRTRYFIPDTGNLAVNRFAQNPKENKAWYILDSNGYAVTGLQTINGQYFYDNE	1361
Qy	1334	GSQVKGGVKNADGTSKYKKEGGELVTNEFFTTDGNVWVYAGANGKTVTGAOVINGOHL	1393
Db	1362	GRQVKGHFVITINQRY--FLDGSQGEIAPSRFVTENNKNWYIYVDGNGKLVKGAQVINGNH	1419
Qy	1394	YFNADGSQVKGGVYKNADGTSKYKNASTGBERLTNEFFTTGDNNWYITGANGKSVTGEVKI	1453
Db	1420	YFNNDYSQVKGAW---ANGRY--YDGDGSAQVSNQFIQAANQWYILNDQGHKVTGLQNT	1474
Qy	1454	GDDTYFPFAKQKQVKQGTVSAGNRIISYYDGSQKRAVSTWIEIQPGVYVYFPKNGLA	1511
Db	1475	NNKYVYFGSNGAQVKGKLLTV-QGKKCYFPAHTGEQVWNRFBAAAGCWYTFYFNSAQOA	1531

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 1.99074 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-41

Perfect score: 116

Sequence: 1 GNYWNGGQWAAASAAAAGRY 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_21.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*
9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	20	7	Add93661 Streptococcus
2	116	100.0	20	9	Adx37284 Streptococcus
3	104	89.7	20	7	Add93639 Streptococcus
4	104	89.7	20	9	Adx37282 Streptococcus
5	104	89.7	431	7	Add93649 Streptococcus
6	104	89.7	431	7	Add93650 Streptococcus
7	104	89.7	431	7	Add93653 Streptococcus
8	104	89.7	431	9	Adx37272 Streptococcus
9	104	89.7	431	9	Adx37273 Streptococcus
10	104	89.7	431	9	Adx37276 Streptococcus
11	104	89.7	431	9	Aeb91500 Microbial
12	104	89.7	432	7	Add93651 Streptococcus
13	104	89.7	432	7	Add93652 Streptococcus
14	104	89.7	432	9	Adx37274 Streptococcus
15	104	89.7	432	9	Adx37275 Streptococcus
16	91	78.4	169	8	Adx37276 Streptococcus
17	91	78.4	169	8	Adx37277 Streptococcus
18	91	78.4	169	8	Adx37278 Streptococcus
19	91	78.4	169	8	Adx37279 Streptococcus
20	90	77.6	486	5	Abp66152 Bifidobacterium
21	89	76.7	132	2	Aaw60944 Streptococcus
22	89	76.7	392	6	Abu02747 S. pneumoniae
23	89	76.7	392	8	Adk47859 Streptococcus
24	89	76.7	392	8	Adt50227 S. pneumoniae

ALIGNMENTS

RESULT 1

Add93661

ID ADD93661 standard; peptide; 20 AA.

XX AC ADD93661;

XX DT 29-JAN-2004 (first entry)

XX DE Streptococcus mutans glucan binding protein-B peptide fragment.

XX KW Glucan binding protein-B; GbpB; vaccine; antitoxins; immunogen.

XX OS Streptococcus mutans.

XX PN WO2003075845-A2.

XX PD 18-SEP-2003.

XX PF 07-MAR-2003; 2003WO-US006962.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

XX Claim 11; Page 10; 49pp; English.

XX The present sequence is that of a peptide fragment of the glucan binding protein-B (GbpB) of Streptococcus mutans strain SK32 ADD93649. The peptide binds to a major histocompatibility complex (MHC) class II protein. It was identified as a potential B cell epitope using a matrix-based algorithm for epitope prediction, which was used to search the primary amino acid sequence of GbpB for known MHC class II binding motifs. The peptide can be used in immunogenic compositions and subunit vaccines for dental caries. These comprise MHC class II protein-binding GbpB peptides covalently linked with peptide subunits (preferably from the catalytic domain) of a glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be prepared synthetically or

Adt50226 S. pneumoniae
Adt50165 S. pneumoniae
Adt94595 Novel S.
Aea58465 Streptococcus
Abp29842 Streptococcus
Abp27864 Streptococcus
Abp25919 Streptococcus
Adt93884 S. pyogenes
Aeb91542 Microbial
Abp25918 Streptococcus
Abp29684 Streptococcus
Adu9524 S. agalactiae
Adv88392 Streptococcus
Adv81808 Streptococcus
Adv79645 Streptococcus
Aeb91642 Microbial
Abp66257 Bifidobacterium
Abp40216 Staphylococcus
Adt97279 Staphylococcus
Aag82294 S. epidermidis
Aag82888 S. epidermidis

CC by recombinant DNA technology. Antibodies raised against MHC class II
 CC binding fragments of GbpB can be used in passive immunisation.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 116; DB 7; Length 20;
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAGRY 20
 |||||
 Db 1 GNYWNGGQWAASAAAAGRY 20

RESULT 2
 ADX37284
 ID ADX37284 standard; peptide; 20 AA.

XX AC ADX37284;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus mutant glucan binding protein B peptide #29.

XX KW immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.

XX OS Streptococcus mutans.

XX PN US2005031633-A1.

XX PD 10-FEB-2005.

XX PF 09-MAR-2004; 2004US-00797821.

XX PR 13-APR-1998; 98US-0081550P.

XX PR 08-JAN-1999; 99US-0115142P.

XX PR 12-APR-1999; 99US-00290049.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PR 07-MAR-2003; 2003US-00383930.

XX PA (SMIT/) SMITH D J.

XX PA (TAUB/) TAUBMAN M A.

XX PI Smith DJ, Taubman MA;

XX PR WPI; 2005-151644/16.

XX PT New composition comprising a fragment of a glucan binding protein-B
 PT (GbpB) that binds to MHC class II protein, and a biocompatible
 PT microparticle, useful for producing an antibody (claimed) for immunizing
 PT mammals against dental caries.

XX PS Claim 4; SEQ ID NO 41; 73pp; English.

XX CC The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus mutans GbpB-derived peptide of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 116; DB 9; Length 20;
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAGRY 20
 |||||
 Db 1 GNYWNGGQWAASAAAAGRY 20

RESULT 3
 ADD93639
 ID ADD93639 standard; peptide; 20 AA.

XX AC ADD93639;

XX DT 29-JAN-2004 (first entry)

XX DE Streptococcus mutans glucan binding protein-B peptide fragment.

XX KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX OS Streptococcus mutans.

XX PN WO2003075845-A2.

XX PD 18-SEP-2003.

XX PF 07-MAR-2003; 2003WO-US006962.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PA (FORS-) FORSYTH INST.

XX PI Smith DJ, Taubman MA;

XX PS WPI; 2003-845091/78.

XX PT Composition useful as vaccines for dental caries comprises a fragment of
 PT a glucan binding protein-B binding to a major histocompatibility complex
 PT class II protein.

XX PS Claim 11; Page 10; 49pp; English.

XX CC The present sequence is that of a peptide comprising amino acid residues
 CC 349-368 of the glucan binding protein-B (GbpB) of Streptococcus mutans
 CC strain SK32 ADD93649. The peptide binds to a major histocompatibility
 CC complex (MHC) class II protein. It was identified as a potential B cell
 CC epitope using a matrix-based algorithm for epitope prediction, which was
 CC used to search the primary amino acid sequence of GbpB for known MHC
 CC class II binding motifs. The peptide can be used in immunogenic
 CC compositions and subunit vaccines for dental caries. These comprise MHC
 CC class II protein-binding GbpB peptides covalently linked with peptide
 CC subunits (preferably from the catalytic domain) of a glucosyltransferase.
 CC The compositions are used in a claimed method of eliciting production of
 CC an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
 CC prepared synthetically or by recombinant DNA technology. Antibodies
 CC raised against MHC class II binding fragments of GbpB can be used in
 CC passive immunisation.

XX SQ Sequence 20 AA;

Query Match 89.7%; Score 104; DB 7; Length 20;
 Best Local Similarity 100.0%; Pred. No. 2e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
 |||||
 Db 1 GNYWNGGQWAASAAAAG 18

RESULT 4
 ADX37262
 ID ADX37262 standard; peptide; 20 AA.

XX AC ADX37262;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus mutant glucan binding protein B peptide #19.

KW immunogenicity; immune stimulation; glucan binding protein-B;
 XX microparticle; major histocompatibility complex; tooth disease.

OS Streptococcus mutans.

PN US2005031633-A1.

XX 10-FEB-2005.

PD 09-MAR-2004; 2004US-00797821.

XX 13-APR-1998; 98US-0081550P.

PR 08-JAN-1999; 99US-0115142P.

PR 12-APR-1999; 99US-00290049.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

PR 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.

PA (TAUB/) TAUBMAN M A.

XX Smith DJ, Taubman MA;

XX WPI; 2005-151644/16.

DR Claim 4; SEQ ID NO 19; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus mutans GbpB-derived peptide of the invention.

XX Sequence 20 AA;

Query Match 89.7%; Score 104; DB 9; Length 20;

Best Local Similarity 100.0%; Pred. No. 2e-06;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GNYWNGGQWAAASAAAAG 18

Db 1 GNYWNGGQWAAASAAAAG 18

RESULT 5

ADD93649

ID ADD93649 standard; protein; 431 AA.

XX ADD93649;

XX 29-JAN-2004 (first entry)

XX Streptococcus mutans glucan binding protein-B.

XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

XX Key Location/Qualifiers

FT Region 6..25

FT /note= "HLA-binding peptide"

FT Region 16..35

FT /note= "HLA-binding peptide"

FT Region 33..52

FT /note= "HLA-binding peptide"

FT Region 37..56

FT /note= "HLA-binding peptide"

FT Region 48..67
 FT /note= "HLA-binding peptide"
 FT Region 52..71
 FT /note= "HLA-binding peptide"
 FT Region 88..107
 FT /note= "HLA-binding peptide"
 FT Region 113..132
 FT /note= "HLA-binding peptide"
 FT Region 117..136
 FT /note= "HLA-binding peptide"
 FT Region 137..156
 FT /note= "HLA-binding peptide"
 FT Region 174..193
 FT /note= "HLA-binding peptide"
 FT Region 194..213
 FT /note= "HLA-binding peptide"
 FT Region 214..233
 FT /note= "HLA-binding peptide"
 FT Region 248..267
 FT /note= "HLA-binding peptide"
 FT Region 289..308
 FT /note= "HLA-binding peptide"
 FT Region 306..325
 FT /note= "HLA-binding peptide"
 FT Region 311..330
 FT /note= "HLA-binding peptide"
 FT Region 349..368
 FT /note= "HLA-binding peptide"
 FT Region 365..384
 FT /note= "HLA-binding peptide"
 FT Region 383..402
 FT /note= "HLA-binding peptide"
 FT Region 403..422
 FT /note= "HLA-binding peptide"

PN WO2003075845-A2.

XX 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US006962.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

PA (FORS-) FORSYTH INST.

PI Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX GENBANK; AY046410.

XX Composition useful as vaccines for dental caries comprises a fragment of
 PT a glucan binding protein-B binding to a major histocompatibility complex
 PT class II protein.

XX Claim 6; Page 7; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding
 CC protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
 CC invention provides immunogenic compositions and vaccines for dental
 CC caries. The compositions comprise major histocompatibility complex (MHC)
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
 CC covalently linked with peptide subunits of a glucosyltransferase. The
 CC compositions are used in a claimed method of eliciting production of an
 CC antibody in a mammal. Dieptopic or multi-epitopic polypeptides can be
 CC prepared synthetically or by recombinant DNA technology. Antibodies
 CC raised against MHC class II binding fragments of GbpB can be used in
 CC passive immunisation.

XX Sequence 431 AA;

Query Match 89.7%; Score 104; DB 7; Length 431;

CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus mutans GbpB protein of the invention.

XX Sequence 431 AA;

Query Match 1 89.7%; Score 104; DB 9; Length 431;
 Best Local Similarity 100.0%; Pred. No. 4.2e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
 |||||
 Db 349 GNYWNGGOWAASAAAAG 366

RESULT 11
 AEB91500
 ID AEB91500 standard; protein; 431 AA.

XX AEB91500;

AC 20-OCT-2005 (first entry)

DT Microbial pathogen adhesin protein sequence, SEQ ID NO:210.

DE algorithm; adhesin; pharmaceutical; vaccine; drug screening;
 KW bordetella pertussis infection; antibacterial; pneumonia;
 KW antiinflammatory; respiratory-gen.; gastric ulcer; antitumor;
 KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.

XX Streptococcus mutans.

OS WO2005076010-A2.

XX 18-AUG-2005.

XX 07-FEB-2005; 2005WO-IN000037.

XX 06-FEB-2004; 2004IN-DE000173.

PR 20-JUL-2004; 2004US-0589227P.

XX (COUL) COUNCIL SCI & IND RES SOUTH AFRICA.

XX Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;

XX WPI; 2005-597835/61.

XX Computational method for identifying adhesin and adhesin like molecules,
 PT comprises computing sequence-based attributes of protein sequences using
 PT neural network software and training an artificial neural network.

XX Claim 16; SEQ ID NO 210; 402pp; English.

XX The present invention relates to a computational method (M1) for
 CC identifying adhesin and adhesin-like proteins, by computing the sequence-
 CC based attributes of protein sequences using five attribute modules of a
 CC neural network software, training an artificial neural network (ANN) for
 CC each of the computed five attributes, and identifying the adhesin and
 CC adhesin-like proteins having probability of being an adhesin (Pad) as
 CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes
 CC encoding adhesin and adhesin-like proteins, having 274 fully defined 162-
 CC 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
 CC genes encoding adhesin and adhesin-like proteins, having 105 fully
 CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated
 CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
 CC pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
 CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:
 CC 280-384) sequences; and a fully connected multilayer feed forward ANN (1)
 CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like
 CC proteins, of therapeutic potential, and identifying and short-listing
 CC proteins for further testing in development of new vaccine formulations
 CC to eliminate diseases caused by various pathogenic organisms. (M1) is

CC useful for identifying putative adhesins that are important in drug
 CC discovery and preventing therapeutics for whooping cough, pneumonia, from
 CC gastric ulcer and urinary tract infections. (M1) identifies adhesins, from
 CC distantly related organisms, and from bacteria belonging to a wide
 CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
 CC unique proteins. The present sequence is a microbial pathogen adhesin
 CC protein sequence.

XX Sequence 431 AA;

Query Match 89.7%; Score 104; DB 9; Length 431;
 Best Local Similarity 100.0%; Pred. No. 4.2e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
 |||||
 Db 349 GNYWNGGOWAASAAAAG 366

RESULT 12
 ADD93651
 ID ADD93651 standard; protein; 432 AA.

XX ADD93651;

XX 29-JAN-2004 (first entry)

XX Streptococcus mutans glucan binding protein-B.

XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

XX WO2003075845-A2.

XX 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US006962.

XX 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman WA;

XX WPI; 2003-845091/78.

XX GENBANK; AY046412.

XX Composition useful as vaccines for dental caries comprises a fragment of
 PT a glucan binding protein-B binding to a major histocompatibility complex
 PT class II protein.

XX Claim 5; Page 8; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding
 CC protein-B (GbpB) of Streptococcus mutans strain 15Jp2. The sequence
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
 CC invention provides immunogenic compositions and vaccines for dental
 CC caries. The compositions comprise major histocompatibility complex (MHC)
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
 CC covalently linked with peptide subunits of a glucosyltransferase. The
 CC compositions are used in a claimed method of eliciting production of an
 CC antibody in a mammal. Dieptopic or multi-epitopic polypeptides can be
 CC prepared synthetically or by recombinant DNA technology. Antibodies
 CC raised against MHC class II binding fragments of GbpB can be used in
 CC passive immunisation.

XX Sequence 432 AA;

Query Match 89.7%; Score 104; DB 7; Length 432;
 Best Local Similarity 100.0%; Pred. No. 4.2e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 GNYWNGGQWAASAAAAG 18
Db      350 GNYWNGGQWAASAAAAG 367

RESULT 13
ADD93652
ID      ADD93652 standard; protein; 432 AA.
XX
AC      ADD93652;
XX
DT      29-JAN-2004 (first entry)
XX
DE      Streptococcus mutans glucan binding protein-B.
XX
KW      Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS      Streptococcus mutans.
XX
PN      WO2003075845-A2.
XX
PD      18-SEP-2003.
XX
PF      07-MAR-2003; 2003WO-US006962.
XX
PR      07-MAR-2002; 2002US-0363209P.
PR      08-AUG-2002; 2002US-0402483P.
XX
PA      (FORS-) FORSYTH INST.
XX
PI      Smith DJ, Taubman MA;
XX
WPI; 2003-845091/78.
DR      GENBANK; AY046413.
XX
CC      Composition useful as vaccines for dental caries comprises a fragment of
PT      a glucan binding protein-B binding to a major histocompatibility complex
PT      class II protein.
XX
PS      Claim 5; Page 8; 49pp; English.
XX
CC      The present sequence is the protein sequence of the glucan binding
CC      protein-B (GbpB) of Streptococcus mutans strain 3SNI. The sequence
CC      includes a number of human leucocyte antigen (HLA)-binding epitopes. The
CC      invention provides immunogenic compositions and vaccines for dental
CC      caries. The compositions comprise major histocompatibility complex (MHC)
CC      class II protein-binding GbpB peptides, especially HLA-binding peptides,
CC      covalently linked with peptide subunits of a glucosyltransferase. The
CC      compositions are used in a claimed method of eliciting production of an
CC      antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
CC      prepared synthetically or by recombinant DNA technology. Antibodies
CC      raised against MHC class II binding fragments of GbpB can be used in
CC      passive immunisation.
XX
SQ      Sequence 432 AA;

Query Match      89.7%; Score 104; DB 7; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GNYWNGGQWAASAAAAG 18
Db      350 GNYWNGGQWAASAAAAG 367

RESULT 14
ADX37274
ID      ADX37274 standard; protein; 432 AA.
XX
AC      ADX37274;
XX
DT      21-APR-2005 (first entry)
XX

Query Match      89.7%; Score 104; DB 9; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GNYWNGGQWAASAAAAG 18
Db      350 GNYWNGGQWAASAAAAG 367

RESULT 15
ADX37275
ID      ADX37275 standard; protein; 432 AA.
XX
AC      ADX37275;
XX
DT      21-APR-2005 (first entry)
XX
DE      Streptococcus mutans glucan binding protein B variant #4.
XX
KW      immunogenicity; immune stimulation; glucan binding protein-B;
KW      microparticle; major histocompatibility complex; tooth disease.
XX
OS      Streptococcus mutans.
XX
PN      US2005031633-A1.
XX
PD      10-FEB-2005.
XX
PF      09-MAR-2004; 2004US-00797821.

```

Streptococcus mutant glucan binding protein B variant #3.
immunogenicity; immune stimulation; glucan binding protein-B;
microparticle; major histocompatibility complex; tooth disease.
Streptococcus mutans.
US2005031633-A1.
10-FEB-2005.
09-MAR-2004; 2004US-00797821.
13-APR-1998; 98US-0081550P.
08-JAN-1999; 99US-0115142P.
12-APR-1999; 99US-00290049.
07-MAR-2002; 2002US-0363209P.
08-AUG-2002; 2002US-0402483P.
07-MAR-2003; 2003US-00383930.
(SMIT/) SMITH D J.
(TAUB/) TAUBMAN M A.
Smith DJ, Taubman MA;
WPI; 2005-151644/16.
New composition comprising a fragment of a glucan binding protein-B
(GbpB) that binds to MHC class II protein, and a biocompatible
microparticle, useful for producing an antibody (claimed) for immunizing
mammals against dental caries.
Claim 3; SEQ ID NO 31; 73pp; English.
The invention relates to a composition comprising a fragment of a glucan
binding protein-B (GbpB) and a biocompatible microparticle, where the
fragment binds to a major histocompatibility complex (MHC) class II
protein. The composition is useful for producing an antibody for
immunizing mammals against dental caries. This sequence corresponds to a
Streptococcus mutans GbpB protein of the invention.
Sequence 432 AA;
Query Match 89.7%; Score 104; DB 9; Length 432;
Best Local Similarity 100.0%; Pred. No. 4.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GNYWNGGQWAASAAAAG 18
Db 350 GNYWNGGQWAASAAAAG 367
RESULT 15
ADX37275
ID ADX37275 standard; protein; 432 AA.
XX
AC ADX37275;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #4.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.

```

XX 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX PI Smith DJ, Taubman MA;
XX DR WPI; 2005-151644/16.
XX PT New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX PS Claim 3; SEQ ID NO 32; 73pp; English.
XX CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX SQ Sequence 432 AA;
XX Query Match 89.7%; Score 104; DB 9; Length 432;
XX Best Local Similarity 100.0%; Pred. No. 4.2e-05;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GNYWNGGQWAAASAAAAG 18
DB 350 GNYWNGGQWAAASAAAAG 367
Search completed: February 11, 2006, 19:15:28
Job time : 3.99074 secs

```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 0.437986 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-41

Perfect score: 116
Sequence: 1 GNYWNGGQWASAAAAGRY 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5 COMB.pep.*
2: /cgn2_6/prodata/1/iaa/6 COMB.pep.*
3: /cgn2_6/prodata/1/iaa/H COMB.pep.*
4: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
5: /cgn2_6/prodata/1/iaa/RE COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	76.7	392	2	US-09-583-110-4374
2	89	76.7	399	2	Sequence 4374, Ap
3	61	52.6	138	2	Sequence 3230, Ap
4	61	52.6	149	2	Sequence 5061, Ap
5	61	52.6	157	2	Sequence 1682, Ap
6	57	49.1	266	2	Sequence 2870, Ap
7	56	48.3	257	2	Sequence 5453, Ap
8	56	48.3	264	2	Sequence 3244, Ap
9	56	48.3	267	2	Sequence 5035, Ap
10	54.5	47.0	626	1	Sequence 4539, Ap
11	54.5	47.0	626	1	Sequence 7, Appli
12	54	46.6	461	1	Sequence 14, Appl
13	52	44.8	322	2	Sequence 2, Appli
14	51	44.0	406	2	Sequence 16924, A
15	50	43.1	113	2	Sequence 20860, A
16	50	43.1	214	2	Sequence 6331, Ap
17	49	42.2	275	2	Sequence 14258, A
18	49	42.2	339	2	Sequence 16149, A
19	49	42.2	445	2	Sequence 26841, A
20	48	41.4	28	1	Sequence 13437, A
21	48	41.4	28	1	Sequence 43, Appl
22	48	41.4	624	1	Sequence 43, Appl
23	48	41.4	651	2	Sequence 9, Appli
24	48	41.4	651	2	Sequence 19, Appl
25	48	41.4	651	2	Sequence 1, Appli
26	48	41.4	651	2	Sequence 1, Appli
27	48	41.4	718	1	Sequence 2, Appli

28	48	41.4	718	1	US-08-317-844B-2	Sequence 2, Appli
29	48	41.4	747	2	US-09-034-177-3	Sequence 3, Appli
30	48	41.4	748	2	US-09-252-991A-32942	Sequence 32942, A
31	47.5	40.9	408	2	US-09-252-991A-31571	Sequence 31571, A
32	47	40.5	146	2	US-08-537-871A-37	Sequence 37, Appl
33	47	40.5	232	2	US-09-252-991A-32746	Sequence 32746, A
34	47	40.5	250	2	US-09-248-796A-22487	Sequence 22487, A
35	46.5	40.1	219	2	US-09-380-015B-2	Sequence 2, Appli
36	46.5	40.1	252	2	US-09-431-887-32	Sequence 32, Appli
37	46.5	40.1	255	1	US-08-242-188-4	Sequence 4, Appli
38	46.5	40.1	255	1	US-08-509-261A-4	Sequence 4, Appli
39	46.5	40.1	255	1	US-08-660-628-10	Sequence 10, Appl
40	46.5	40.1	255	1	US-08-692-892-4	Sequence 4, Appli
41	46.5	40.1	255	1	US-08-713-939A-4	Sequence 4, Appli
42	46.5	40.1	255	1	US-08-868-162A-24	Sequence 24, Appl
43	46.5	40.1	255	2	US-09-031-168-10	Sequence 10, Appl
44	46.5	40.1	255	2	US-09-036-579-4	Sequence 4, Appli
45	46.5	40.1	255	2	US-09-550-374-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-583-110-4374
; Sequence 4374, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4374
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4374

Query Match 76.7%; Score 89; DB 2; Length 392;
Best Local Similarity 83.3%; Pred. No. 0.0011;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 GNYWNGGQWASAAAAG 18
Db 304 GDYWGNGAQWATSAAGA 321

RESULT 2
US-09-107-433-3230
; Sequence 3230, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354

<p>COMPUTER READABLE FORM: MEDIUM TYPE: CD/ROM ISO9660 COMPUTER: <Unknown> OPERATING SYSTEM: <Unknown> SOFTWARE: <Unknown> CURRENT APPLICATION DATA: APPLICATION NUMBER: US/09/107,433 FILING DATE: 30-Jun-1998 PRIOR APPLICATION DATA: APPLICATION NUMBER: 60/ 085131 FILING DATE: May 12, 1998 APPLICATION NUMBER: 60/051553 FILING DATE: July 2, 1997 ATTORNEY/AGENT INFORMATION: NAME: Ariniello, Pamela Deneke REGISTRATION NUMBER: 40,489 REFERENCE/DOCKET NUMBER: GTC-011 TELECOMMUNICATION INFORMATION: TELEPHONE: (781)893-5007 TELEFAX: (781)893-8277 INFORMATION FOR SEQ ID NO: 3230: SEQUENCE CHARACTERISTICS: LENGTH: 399 amino acids TYPE: amino acid TOPOLOGY: linear MOLECULE TYPE: protein HYPOTHETICAL: YES ORIGINAL SOURCE: ORGANISM: Streptococcus pneumoniae FEATURE: NAME/KEY: misc feature LOCATION: (B) LOCATION 1...399 SEQUENCE DESCRIPTION: SEQ ID NO: 3230: US-09-107-433-3230</p>	<p>Query Match 1 76.7%; Score 89; DB 2; Length 399; Best Local Similarity 83.3%; Pred. No. 0.0011; Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;</p> <p>QY 1 GNYWNGGQWAAASAAAAG 18 : : : Db 311 GDYWGNGAQWATSAAAAAG 328</p> <p>RESULT 3 US-09-134-001C-5061 Sequence 5061, Application US/09134001C Patent No. 6380370 GENERAL INFORMATION: APPLICANT: Lynn Doucette-Stamm et al TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS FILE REFERENCE: GTC-007 CURRENT APPLICATION NUMBER: US/09/134,001C CURRENT FILING DATE: 1998-08-13 PRIOR APPLICATION NUMBER: US 60/064,964 PRIOR FILING DATE: 1997-11-08 PRIOR APPLICATION NUMBER: US 60/055,779 PRIOR FILING DATE: 1997-08-14 NUMBER OF SEQ ID NOS: 5674 SEQ ID NO 5061 LENGTH: 138 TYPE: PRT ORGANISM: Staphylococcus epidermidis US-09-134-001C-5061</p> <p>Query Match 52.6%; Score 61; DB 2; Length 138; Best Local Similarity 55.6%; Pred. No. 1.2; Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;</p> <p>QY 1 GNYWNGGQWAAASAAAAG 18 : : : Db 82 GSTWGNANSWATAQAAG 99</p>
<p>COMPUTER READABLE FORM: MEDIUM TYPE: CD/ROM ISO9660 COMPUTER: <Unknown> OPERATING SYSTEM: <Unknown> SOFTWARE: <Unknown> CURRENT APPLICATION DATA: APPLICATION NUMBER: US/09/107,433 FILING DATE: 30-Jun-1998 PRIOR APPLICATION DATA: APPLICATION NUMBER: 60/ 085131 FILING DATE: May 12, 1998 APPLICATION NUMBER: 60/051553 FILING DATE: July 2, 1997 ATTORNEY/AGENT INFORMATION: NAME: Ariniello, Pamela Deneke REGISTRATION NUMBER: 40,489 REFERENCE/DOCKET NUMBER: GTC-011 TELECOMMUNICATION INFORMATION: TELEPHONE: (781)893-5007 TELEFAX: (781)893-8277 INFORMATION FOR SEQ ID NO: 3230: SEQUENCE CHARACTERISTICS: LENGTH: 399 amino acids TYPE: amino acid TOPOLOGY: linear MOLECULE TYPE: protein HYPOTHETICAL: YES ORIGINAL SOURCE: ORGANISM: Streptococcus pneumoniae FEATURE: NAME/KEY: misc feature LOCATION: (B) LOCATION 1...399 SEQUENCE DESCRIPTION: SEQ ID NO: 3230: US-09-107-433-3230</p>	<p>Query Match 1 76.7%; Score 89; DB 2; Length 399; Best Local Similarity 83.3%; Pred. No. 0.0011; Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;</p> <p>QY 1 GNYWNGGQWAAASAAAAG 18 : : : Db 311 GDYWGNGAQWATSAAAAAG 328</p> <p>RESULT 3 US-09-134-001C-5061 Sequence 5061, Application US/09134001C Patent No. 6380370 GENERAL INFORMATION: APPLICANT: Lynn Doucette-Stamm et al TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS FILE REFERENCE: GTC-007 CURRENT APPLICATION NUMBER: US/09/134,001C CURRENT FILING DATE: 1998-08-13 PRIOR APPLICATION NUMBER: US 60/064,964 PRIOR FILING DATE: 1997-11-08 PRIOR APPLICATION NUMBER: US 60/055,779 PRIOR FILING DATE: 1997-08-14 NUMBER OF SEQ ID NOS: 5674 SEQ ID NO 5061 LENGTH: 138 TYPE: PRT ORGANISM: Staphylococcus epidermidis US-09-134-001C-5061</p> <p>Query Match 52.6%; Score 61; DB 2; Length 138; Best Local Similarity 55.6%; Pred. No. 1.2; Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;</p> <p>QY 1 GNYWNGGQWAAASAAAAG 18 : : : Db 82 GSTWGNANSWATAQAAG 99</p>

;
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5453
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5453

Query Match 49.1%; Score 57; DB 2; Length 266;
Best Local Similarity 55.6%; Pred. No. 7.6;
Matches 10; Conservative 1; Mismatches 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
|: ||| ||| ||| :|| :||
Db 180 GSLWGNASNWAYSANQAG 197

RESULT 7
US-09-710-279-3244
; Sequence 3244, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-3244

Query Match 48.3%; Score 56; DB 2; Length 257;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
|: ||| ||| ||| :|| :||
Db 174 GSTWGNANNWANAARSG 191

RESULT 8
US-09-134-001C-5035
; Sequence 5035, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5035

;
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5035

Query Match 48.3%; Score 56; DB 2; Length 264;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
|: ||| ||| ||| :|| :||
Db 181 GSTWGNANNWANAARSG 198

RESULT 9
US-09-134-001C-4539
; Sequence 4539, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4539
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4539

Query Match 48.3%; Score 56; DB 2; Length 267;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
|: ||| ||| ||| :|| :||
Db 184 GSTWGNANNWANAARSG 201

RESULT 10
US-08-596-300A-7
; Sequence 7, Application US/08596300A
; Patent No. 5834191
; GENERAL INFORMATION:
; APPLICANT: Production of Heterologous Peptides
; TITLE OF INVENTION: 14
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596,300A
; FILING DATE: 13-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-001

TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 626 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE:
US-08-596-300A-7

Query Match 47.0%; Score 54.5; DB 1; Length 626;
Best Local Similarity 57.9%; Pred. No. 37;
Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 YWG-NGGQWAAASAAAAGRY 20
Db 323 FWGNSGRTAGKAAAVGRY 341

RESULT 11
US-08-596-300A-14
Sequence 14, Application US/08596300A
Patent No. 5834191
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Production of Heterologous Peptides
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,300A
FILING DATE: 13-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1321-1-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 626 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE:
US-08-596-300A-14

Query Match 47.0%; Score 54.5; DB 1; Length 626;
Best Local Similarity 57.9%; Pred. No. 37;
Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 YWG-NGGQWAAASAAAAGRY 20
: | | | | | | | | | | | |

Db 323 FWGNSGRTAGKAAAVGRY 341

RESULT 12
US-08-186-222-2
Sequence 22, Application US/08186222
Patent No. 5559007
GENERAL INFORMATION:
APPLICANT: Suri, Bruno
APPLICANT: Schmitz, Albert
TITLE OF INVENTION: Bacterial Vectors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/186,222
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US 07/672,205
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: GB 9006400.7
FILING DATE: 22-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Villamizar, JoAnn
REGISTRATION NUMBER: 30,598
REFERENCE/DOCKET NUMBER: 4-17994/A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 785-7121
TELEFAX: (914) 347-5769
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-186-222-2

Query Match 46.6%; Score 54; DB 1; Length 461;
Best Local Similarity 64.3%; Pred. No. 32;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GNGGQWAAASAAAAG 18
: | | | | | | | | | | | |

Db 376 GNGGQWASNGPQAG 389

RESULT 13
US-09-248-796A-16924
Sequence 16924, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208

```
; SEQ ID NO 16924
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16924

Query Match      44.8%; Score 52; DB 2; Length 322;
Best Local Similarity 72.7%; Pred. No. 39;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GNYWNGGGQWA 11
Db      145 GHYSGNGGQWS 155

RESULT 14
US-09-248-796A-20860
; Sequence 20860, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20860
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20860

Query Match      44.0%; Score 51; DB 2; Length 406;
Best Local Similarity 47.4%; Pred. No. 66;
Matches 9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY      1 GNYWNGGGQWAASAAAGR 19
Db      377 GNYRGNGNWRGNSNRGR 395

RESULT 15
US-09-513-999C-6331
; Sequence 6331, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 6331
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 24
; OTHER INFORMATION: Xaa=Ala or Ser
; FEATURE:
; NAME/KEY: UNSURE
```

```
; LOCATION: 82
; OTHER INFORMATION: Xaa=Ala or Pro
US-09-513-999C-6331

Query Match      43.1%; Score 50; DB 2; Length 113;
Best Local Similarity 55.0%; Pred. No. 24;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      1 GNYWNGGGQWAASAAAGRY 20
Db      26 GNYFGFTHSGAARAAQAQY 45

Search completed: February 11, 2006, 19:46:19
Job time : 1.43789 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 1.79012 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-41
Perfect score: 116
Sequence: 1 GNYWNGGOWAASAAAAGRY 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:*
1: /cgn2_6/prodata1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/prodata1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/prodata1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	100.0	20	4	US-10-383-930-41
2	116	100.0	20	5	US-10-797-821-41
3	104	89.7	20	4	US-10-383-930-19
4	104	89.7	20	5	US-10-797-821-19
5	104	89.7	431	4	US-10-383-930-29
6	104	89.7	431	4	US-10-383-930-30
7	104	89.7	431	4	US-10-383-930-33
8	104	89.7	431	5	US-10-797-821-29
9	104	89.7	431	5	US-10-797-821-30
10	104	89.7	431	5	US-10-797-821-33
11	104	89.7	432	4	US-10-383-930-31
12	104	89.7	432	4	US-10-383-930-32
13	104	89.7	432	5	US-10-797-821-31
14	104	89.7	432	5	US-10-797-821-32
15	89	76.7	392	5	US-10-472-928-4652
16	89	76.7	399	5	US-10-617-320-3230
17	84	72.4	398	5	US-10-474-792-600
18	61	52.6	138	4	US-10-724-972A-6574
19	61	52.6	157	5	US-10-470-048B-361
20	61	52.6	166	5	US-10-470-048B-73
21	59	50.9	267	5	US-10-470-048B-74
22	58	50.0	261	5	US-10-470-048B-65
23	57	49.1	266	4	US-10-724-972A-7113
24	57	49.1	360	4	US-10-724-972A-7511
25	56	48.3	77	4	US-10-106-698-6091
26	56	48.3	257	4	US-10-470-048B-362
27	56	48.3	264	4	US-10-724-972A-6539

28	56	48.3	267	4	US-10-724-972A-5110	Sequence 5110, Ap
29	54.5	47.0	647	4	US-10-369-493-3109	Sequence 3109, Ap
30	54	46.6	277	4	US-10-437-963-113400	Sequence 113400, Ap
31	53	45.7	108	6	US-11-097-143-2586	Sequence 2586, Ap
32	52	44.8	101	5	US-10-450-763-57503	Sequence 57503, A
33	52	44.8	363	4	US-10-437-963-170917	Sequence 170917, A
34	52	44.8	1968	4	US-10-369-493-6942	Sequence 6942, Ap
35	52	44.8	1968	4	US-10-369-493-6943	Sequence 6943, Ap
36	51	44.0	113	4	US-10-425-115-218118	Sequence 218118, A
37	51	44.0	375	3	US-09-917-378-3	Sequence 3, Appli
38	51	44.0	375	3	US-09-917-378-6	Sequence 6, Appli
39	51	44.0	762	3	US-09-917-378-1	Sequence 1, Appli
40	51	44.0	1230	4	US-10-437-963-120863	Sequence 120863, A
41	51	44.0	2546	5	US-10-732-923-20517	Sequence 20517, A
42	50	43.1	102	4	US-10-425-115-268735	Sequence 268735, A
43	50	43.1	119	5	US-10-450-763-44127	Sequence 44127, A
44	50	43.1	173	4	US-10-425-115-366795	Sequence 366795, A
45	50	43.1	193	4	US-10-437-963-137412	Sequence 137412, A

ALIGNMENTS

RESULT 1

US-10-383-930-41
; Sequence 41, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-41

Query Match 100.0%; Score 116; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAGRY 20
|||||
Db 1 GNYWNGGOWAASAAAAGRY 20

RESULT 2

US-10-797-821-41
; Sequence 41, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049

```
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: GbpB peptide
US-10-797-821-41
```

```
Query Match      100.0%; Score 116; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GNYWNGGQWAASAAAGRY 20
   |||||
Db 1 GNYWNGGQWAASAAAGRY 20
   |||||
```

RESULT 3

```
US-10-383-930-19
; Sequence 19, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-19
```

```
Query Match      89.7%; Score 104; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.7e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GNYWNGGQWAASAAAG 18
   |||||
Db 1 GNYWNGGQWAASAAAG 18
   |||||
```

RESULT 4

```
US-10-797-821-19
; Sequence 19, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
```

```
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: GbpB peptide
US-10-797-821-19
```

```
Query Match      89.7%; Score 104; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.7e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GNYWNGGQWAASAAAG 18
   |||||
Db 1 GNYWNGGQWAASAAAG 18
   |||||
```

RESULT 5

```
US-10-383-930-29
; Sequence 29, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29
```

```
Query Match      89.7%; Score 104; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GNYWNGGQWAASAAAG 18
   |||||
Db 349 GNYWNGGQWAASAAAG 366
   |||||
```

RESULT 6

```
US-10-383-930-30
; Sequence 30, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
```



```

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PR1
; ORGANISM: Streptococcus mutans
US-10-383-930-30

```

Query Match 89.7%; Score 104; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels

Qy 1 GNTWNGGQWAAASAAAAG 18
|||
349 GNTWNGGQWAAASAAAAG 366

RESULT. T 7

US-10-383-930-33
; Sequence 33, Application US/10383930
; Publication No. US20040127400A1

```

: GENERAL INFORMATION:
: APPLICANT: Smith, Daniel J
: APPLICANT: Taubman, Martin A
: TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
: FILE REFERENCE: 25669-018
: CURRENT APPLICATION NUMBER: US/10/383,930
: CURRENT FILING DATE: 2003-03-07
: PRIOR APPLICATION NUMBER: 60/402,483
: PRIOR FILING DATE: 2002-08-08
: PRIOR APPLICATION NUMBER: 60/363,209
: PRIOR FILING DATE: 2002-03-07
: NUMBER OF SEQ ID NOS: 41

```

```

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431

```

; TYPE: PRT
 ; ORGANISM: Streptococcus mutans
 US-10-383-930-33

Query Match 89.7%; Score 104; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels

Qy 1 GNYWNGGQWAAASAAAAG 18
Db 349 GNYWNGGQWAAASAAAAG 366

RESULT 8

US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
: GENERAL INFORMATION.

; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.

;
 ;
 ; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
 ;
 ; FILE REFERENCE: 25669-020
 ; CURRENT APPLICATION NUMBER: US/10/797,821
 ;

; CURRENT FILING DATE: 2004-03-09
 ; PRIOR APPLICATION NUMBER: 10/383,930
 ; PRIOR FILING DATE: 2003-03-07

```

1 PRIOR APPLICATION NUMBER: 60/363,209
2 PRIOR FILING DATE: 2002-03-07
3 PRIOR APPLICATION NUMBER: 60/402,483
4 PRIOR FILING DATE: 2002-08-08
5 PRIOR APPLICATION NUMBER: 09/290,049
6 PRIOR FILING DATE: 1999-04-12
7 PRIOR APPLICATION NUMBER: 60/081,550
8 PRIOR FILING DATE: 1998-04-13
9 PRIOR APPLICATION NUMBER: 60/115,142
10 PRIOR FILING DATE: 1999-01-08
11 NUMBER OF SEQ ID NOS: 45
12 SOFTWARE: PatentIn version 3.2

```

```

; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: St
US-10-797-821-29

```

Query Match 89.7%; Score 104; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels

Qy 1 GNTWNGGQWAAASAAAAG 18
Db 349 GNTWNGGQWAAASAAAAG 366

RESULT 9

RESULTS 9
IIS-10-797-821-30

US-10-797-821-30
; Sequence 30, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:

APPLICANT: Smith, Daniel J.
APPLICANT: Taubman, Martin A.
TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
FILE REFERENCE: 25669-020

; CURRENT APPLICATION NUMBER: US/10/797,821
 ;
 ; CURRENT FILING DATE: 2004-03-09

```

P R I O R   A P P L I C A T I O N   N U M B E R :   6 0 / 3 8 3 , 9 3 0
P R I O R   F I L I N G   D A T E :   2 0 0 3 - 0 3 - 0 7
P R I O R   A P P L I C A T I O N   N U M B E R :   6 0 / 3 6 3 , 2 0 9
P R I O R   F I L I N G   D A T E :   2 0 0 2 - 0 3 - 0 7
P R I O R   A P P L I C A T I O N   N U M B E R :   6 0 / 4 0 2 , 4 5 3
P R I O R   F I L I N G   D A T E :   2 0 0 2 - 0 8 - 0 8
P R I O R   A P P L I C A T I O N   N U M B E R :   0 9 / 2 9 0 , 0 4 9
P R I O R   F I L I N G   D A T E :   1 9 9 9 - 0 4 - 1 2
P R I O R   A P P L I C A T I O N   N U M B E R :   6 0 / 0 8 1 , 5 5 0
P R I O R   F I L I N G   D A T E :   1 9 9 8 - 0 4 - 1 3
P R I O R   A P P L I C A T I O N   N U M B E R :   6 0 / 1 1 5 , 1 4 2
P R I O R   F I L I N G   D A T E :   1 9 9 9 - 0 1 - 0 8
N U M B E R   O F   S E Q   I D   N O S :   4 5
S O F T W A R E   P A T E N T   I N   V E R S I O N   3 . 2

```

```

; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT

```

ORGANISM: Streptococcus mutans
US-10-797-821-30

Query Match 89.7%; Score 104; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels

Qy 1 GNYWNGGQWAAASAAAAG 18
Db 349 GNYWNGGQWAAASAAAAG 366

RESULT 10

US-10-797-821-33

US 10 797 041-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin

```

; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
;
; FILE REFERENCE: 25669-020
;
; CURRENT APPLICATION NUMBER: US/10/797,821

```

; CURRENT FILING DATE: 2004-03-09
 ; PRIOR APPLICATION NUMBER: 10/383,930

; PRIOR FILING DATE: 2003-03-07
 ; PRIOR APPLICATION NUMBER: 60/363,209
 ; PRIOR FILING DATE: 2003-03-07
 ; PRIOR APPLICATION NUMBER: 60/402,483
 ; PRIOR FILING DATE: 2003-08-08
 ; PRIOR APPLICATION NUMBER: 09/290,049

```

; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33

```

```

Query Match      89.7%; Score 104; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GNYWNGGQWAAASAAAAG 18
        |||||
Db      349 GNYWNGGQWAAASAAAAG 366

```

```

RESULT 11
US-10-383-930-31
; Sequence 31, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-31

```

```

Query Match      89.7%; Score 104; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GNYWNGGQWAAASAAAAG 18
        |||||
Db      350 GNYWNGGQWAAASAAAAG 367

```

```

RESULT 12
US-10-383-930-32
; Sequence 32, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432

```

```

; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-32

```

```

Query Match      89.7%; Score 104; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GNYWNGGQWAAASAAAAG 18
        |||||
Db      350 GNYWNGGQWAAASAAAAG 367

```

```

RESULT 13
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31

```

```

Query Match      89.7%; Score 104; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GNYWNGGQWAAASAAAAG 18
        |||||
Db      350 GNYWNGGQWAAASAAAAG 367

```

```

RESULT 14
US-10-797-821-32
; Sequence 32, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550

```

```

; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-32

```

Query Match 89.7%; Score 104; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 15
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: Seqwin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: secreted 45 kd protein (usp45)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Similar to strain R6 sequence 15904062 (O.E+01)
US-10-472-928-4652

```

Query Match 76.7%; Score 89; DB 5; Length 392;
Best Local Similarity 83.3%; Pred. No. 0.0043;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Search completed: February 11, 2006, 20:57:29
Job time : 2.79012 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - proteip search, using sw model

Run on: February 11, 2006, 20:42:47 ; Search time 0.113812 Seconds
(without alignments)
2306.008 Million cell updates/sec

Title: US-10-797-821-41

Perfect score: 116

Sequence: 1 GNYWNGGQWAAASAAAAGRY 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 1312538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New:**

- 1: /cgn2_6/prodata1/pubpaa/US08_NEW_PUB.pep:**
- 2: /cgn2_6/prodata1/pubpaa/US06_NEW_PUB.pep:**
- 3: /cgn2_6/prodata1/pubpaa/US07_NEW_PUB.pep:**
- 4: /cgn2_6/prodata1/pubpaa/PCT_NEW_PUB.pep:**
- 5: /cgn2_6/prodata1/pubpaa/US09_NEW_PUB.pep:**
- 6: /cgn2_6/prodata1/pubpaa/US10_NEW_PUB.pep:**
- 7: /cgn2_6/prodata1/pubpaa/US11_NEW_PUB.pep:**
- 8: /cgn2_6/prodata1/pubpaa/US60_NEW_PUB.pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	89.7	431	7	US-11-052-554A-210
2	84	72.4	398	7	US-11-052-554A-252
3	78	67.2	211	7	US-11-052-554A-352
4	61	52.6	149	6	US-10-793-626-1682
5	61	52.6	157	6	US-10-793-626-2870
6	61	52.6	544	7	US-11-052-554A-358
7	56	48.3	257	6	US-10-793-626-3244
8	49	42.2	224	6	US-10-857-780-25
9	48	41.4	279	6	US-10-517-939-194
10	46.5	40.1	219	6	US-10-867-589-2
11	46	39.7	210	7	US-11-177-509-34
12	46	39.7	211	7	US-11-089-551A-19
13	46	39.7	235	6	US-10-793-626-948
14	46	39.7	283	6	US-10-517-939-180
15	46	39.7	365	7	US-11-052-554A-223
16	46	39.7	618	7	US-11-052-554A-150
17	45	38.8	128	7	US-11-173-071-12
18	45	38.8	128	7	US-11-173-071-14
19	45	38.8	143	6	US-10-793-626-112
20	45	38.8	143	6	US-10-793-626-1940
21	45	38.8	212	7	US-11-089-551A-21
22	45	38.8	294	7	US-11-052-554A-71
23	45	38.8	388	7	US-11-130-821-1
24	45	38.8	1538	7	US-11-052-554A-146
25	44.5	38.4	123	6	US-10-925-366A-217

26	44	37.9	141	6	US-10-467-657-748	Sequence 748, App
27	44	37.9	249	7	US-11-054-515-974	Sequence 974, App
28	44	37.9	434	7	US-11-052-554A-167	Sequence 167, App
29	44	37.9	606	7	US-11-052-554A-163	Sequence 163, App
30	44	37.9	1288	7	US-11-052-554A-93	Sequence 93, Appl
31	43.5	37.5	254	7	US-11-054-515-1139	Sequence 1139, Ap
32	43	37.1	270	6	US-10-485-517-413	Sequence 413, App
33	43	37.1	330	6	US-10-485-517-415	Sequence 415, App
34	43	37.1	1381	7	US-11-052-554A-138	Sequence 138, App
35	43	37.1	1901	7	US-11-052-554A-135	Sequence 135, App
36	42.5	36.6	1660	7	US-11-052-554A-137	Sequence 137, App
37	42	36.2	101	7	US-11-110-424-1	Sequence 1, Appl1
38	42	36.2	101	7	US-11-110-424-2	Sequence 2, Appl1
39	42	36.2	101	7	US-11-110-424-3	Sequence 3, Appl1
40	42	36.2	210	7	US-11-177-509-35	Sequence 35, Appl
41	42	36.2	210	7	US-11-177-509-36	Sequence 36, Appl
42	42	36.2	210	7	US-11-177-509-37	Sequence 37, Appl
43	42	36.2	210	7	US-11-177-509-38	Sequence 38, Appl
44	42	36.2	210	7	US-11-177-509-39	Sequence 39, Appl
45	42	36.2	210	7	US-11-177-509-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-11-052-554A-210
; Sequence 210, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 210
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-210

Query Match 89.7%; Score 104; DB 7; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAAG 18
DB 349 GNYWNGGQWAAASAAAAG 366

RESULT 2
US-11-052-554A-252
; Sequence 252, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 252
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-252

Query Match 72.4%; Score 84; DB 7; Length 398;
Best Local Similarity 83.3%; Pred. No. 0.00039;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
|:|||||:|:|
Db 315 GNNWNGGOWAYSQAAG 332

RESULT 3

US-11-052-554A-352
; Sequence 352, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 352
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-352

Query Match 67.2%; Score 78; DB 7; Length 211;
Best Local Similarity 66.7%; Pred. No. 0.0013;
Matches 12; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
|:|||||:|:|
Db 122 GDWNGGOWASSAAG 139

RESULT 4

US-10-793-626-1682
; Sequence 1682, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1682
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-1682

Query Match 52.6%; Score 61; DB 6; Length 149;
Best Local Similarity 55.6%; Pred. No. 0.15;

Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 1 GNYWNGGOWAASAAAAG 18
|:|||||:|:|
Db 72 GSTWGNANSWATAQAAG 89

RESULT 5

US-10-793-626-2870
; Sequence 2870, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2870
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-2870

Query Match 52.6%; Score 61; DB 6; Length 157;
Best Local Similarity 55.6%; Pred. No. 0.16;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
|:|||||:|:|
Db 72 GSTWGNANSWATAQAAG 89

RESULT 6

US-11-052-554A-358
; Sequence 358, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 358
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-358

Query Match 52.6%; Score 61; DB 7; Length 544;
Best Local Similarity 58.8%; Pred. No. 0.46;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 NYWNGGOWAASAAAAG 18
|:|||||:|:|
Db 457 NWLNGGOWASTAVKVG 473

RESULT 7

US-10-793-626-3244

; Sequence 3244, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3244

Query Match 48.3%; Score 56; DB 6; Length 257;
Best Local Similarity 50.0%; Pred. No. 1.1;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 GNYWGGGQWAAASAAAG 18
|: ||| ||: ||: |
Db 174 GSTWGNANNWANAARSQ 191

RESULT 8
US-10-857-780-25
; Sequence 25, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-857-780-25

Query Match 42.2%; Score 49; DB 6; Length 224;
Best Local Similarity 58.3%; Pred. No. 7.4;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 WGGGQWAAASAA 15
||| ||: ||: |
Db 199 WGGGQWAAATVA 210

RESULT 9
US-10-517-939-194
; Sequence 194, Application US/10517939
; Publication No. US20060003433A1

; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esteghlalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)...(21)
US-10-517-939-194

Query Match 41.4%; Score 48; DB 6; Length 279;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GNYWNGGQW 10
|: ||| ||: |
Db 101 GFYWGNGGKY 110

RESULT 10
US-10-867-589-2
; Sequence 2, Application US/10867589
; Publication No. US20060025575A1
; GENERAL INFORMATION:
; APPLICANT: Carsten Korth
; TITLE OF INVENTION: Immunological Detection of Prions
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kanton Zuerich vertreten durch die Erziehungsdirektion
; STREET: Walchtor
; CITY: Zuerich
; STATE: Zuerich
; COUNTRY: Switzerland
; ZIP: CH-8090
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/867,589
; FILING DATE: 15-Jun-2004
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/380,015
; FILING DATE: 23-Aug-1999
; APPLICATION NUMBER: EP 97102837.8
; FILING DATE: 21-FEB-1997
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-867-589-2

Query Match      40.1%; Score 46.5; DB 6; Length 219;
Best Local Similarity 39.4%; Pred. No. 15;
Matches 13; Conservative 0; Mismatches 5; Indels 15; Gaps 2;

QY 1 GNYWNGG---GQW-----AASAAAAG 18
DB 75 GGGWGGGGTHGQWNKPSKPTNKLKHVAGAAAAG 107

RESULT 11
US-11-177-509-34
; Sequence 34, Application US/11177509
; Publication No. US20060018918A1
; GENERAL INFORMATION:
; APPLICANT: CHANG, ROWEN J.Y.
; TITLE OF INVENTION: PRODUCTION OF STABILIZED CONFORMATIONAL ISOMERS OF
; FILE REFERENCE: UTHH 1006
; CURRENT APPLICATION NUMBER: US/11/177,509
; CURRENT FILING DATE: 2005-07-08
; PRIOR APPLICATION NUMBER: 10/210,862
; PRIOR FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: 10/025,976
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/258,576
; PRIOR FILING DATE: 2000-12-27
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 34
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-177-509-34

Query Match      39.7%; Score 46; DB 7; Length 210;
Best Local Similarity 38.2%; Pred. No. 17;
Matches 13; Conservative 0; Mismatches 5; Indels 16; Gaps 2;

QY 1 GNYWNGG---GQW-----AASAAAAG 18
DB 64 GGGWGGGGTHGQWNKPSKPTNKLKHVAGAAAAG 97

RESULT 12
US-11-089-551A-19
; Sequence 19, Application US/11089551A
; Publication No. US20050266242A1
; GENERAL INFORMATION:
; APPLICANT: Lindquist et al.
; TITLE OF INVENTION: ELECTRICAL CONDUCTORS AND DEVICES FROM PRION-LIKE PROTEINS
; FILE REFERENCE: 30554/40025A
; CURRENT APPLICATION NUMBER: US/11/089,551A
; CURRENT FILING DATE: 2005-03-24
; PRIOR APPLICATION NUMBER: US 60/559,286
; PRIOR FILING DATE: 2004-03-31
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 211
; TYPE: PRT
; ORGANISM: MOUSE
US-11-089-551A-19

Query Match      39.7%; Score 46; DB 7; Length 211;
Best Local Similarity 38.2%; Pred. No. 17;
```

```

Matches 13; Conservative 0; Mismatches 5; Indels 16; Gaps 2;

QY 1 GNYWNGG---GQW-----AASAAAAG 18
DB 65 GGGWGGGGTHGQWNKPSKPTNKLKHVAGAAAAG 98

RESULT 13
US-10-793-626-948
; Sequence 948, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 948
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-948

Query Match      39.7%; Score 46; DB 6; Length 235;
Best Local Similarity 47.1%; Pred. No. 19;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 NYWNGGQWAAASAAAAG 18
DB 72 NFTSGSEWSMSYAVAG 88

RESULT 14
US-10-517-939-180
; Sequence 180, Application US/10517939
; Publication No. US20060003433A1
; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esteghalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)...(19)
US-10-517-939-180
```

```

Query Match      39.7%; Score 46; DB 6; Length 283;
Best Local Similarity 70.0%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GNYWNGGQW 10
Db      104 GFWWNGGQW 113
      | : ||||| :
      | : ||||| :

RESULT 15
US-11-052-554A-223
; Sequence 223, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 223
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-223

Query Match      39.7%; Score 46; DB 7; Length 365;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 YWNGGQW 10
Db      278 YWNGGQW 285
      | : ||||| :
      | : ||||| :

```

Search completed: February 11, 2006, 20:59:02
Job time : 1.11381 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 0.412809 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-41

Perfect score: 116

Sequence: 1 GNYWNGGQWAAASAAAAGRY 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	76.7	392	2 G95258	secreted 45 kd pro
2	89	76.7	392	2 B98124	general stress pro
3	61	52.6	166	2 C90029	hypothetical prote
4	59	50.9	267	2 F90028	hypothetical prote
5	58	50.0	255	2 G90061	hypothetical prote
6	57	49.1	358	2 H36891	transfer complex p
7	57	49.1	475	2 C86863	N-acetylmuramoyl-l
8	54.5	47.0	626	2 T49625	glucan 1,4-alpha-g
9	54.5	47.0	626	2 S36364	glucan 1,4-alpha-g
10	54	46.6	355	2 E95843	hypothetical prote
11	54	46.6	461	2 JN0097	secreted 45K prote
12	53	45.7	161	2 E85354	PR-1-like protein
13	53	45.7	456	2 E85903	hypothetical prote
14	52.5	45.3	71	2 S45608	light-harvesting p
15	52	44.8	732	2 D64754	probable oxidoredu
16	52	44.8	732	2 E85518	hypothetical prote
17	52	44.8	732	2 B90668	hypothetical prote
18	52	44.8	1968	1 S05697	myosin heavy chain
19	51	44.0	627	2 F95867	conserved hypothet
20	50	43.1	279	2 D89848	conserved hypothet
21	50	43.1	460	2 E85518	hypothetical prote
22	50	43.1	611	2 S60040	alpha-amylase (EC
23	50	43.1	1052	2 T14343	zinc finger RNA bi
24	49.5	42.7	387	2 T52451	endopeptidase Clp
25	49.5	42.7	2639	2 T31328	fibroin - Chinese
26	49	42.2	187	2 G83047	hypothetical prote
27	49	42.2	205	2 G81055	guanylate kinase N
28	49	42.2	283	2 B82860	replication protei
29	49	42.2	502	2 B70845	probable sygar tra

30	48	41.4	257	2 A27872	outer membrane pro
31	48	41.4	562	2 B70953	hypothetical glyci
32	48	41.4	624	2 JC6023	poly(3-hydroxyvalk
33	48	41.4	718	2 A36068	major ampullate fi
34	47	40.5	59	2 T43106	hypothetical prote
35	47	40.5	145	2 A87574	hypothetical prote
36	47	40.5	435	2 T15143	hypothetical prote
37	47	40.5	1246	2 G90887	cryptic nitrate re
38	47	40.5	1246	2 B85730	cryptic nitrate re
39	46.5	40.1	141	2 I47178	Ig heavy chain var
40	46.5	40.1	252	2 JC6175	prion protein - ra
41	46.5	40.1	256	2 JU0268	major prion protei
42	46.5	40.1	256	2 S37149	prion protein - go
43	46.5	40.1	256	2 A54281	major prion protei
44	46.5	40.1	264	2 A54330	major prion protei
45	46.5	40.1	264	2 S37137	prion protein - gr

ALIGNMENTS

RESULT 1

G95258

secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004

C:Accession: G95258

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei

non, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap

neon, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: G95258

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <KUR>

A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMV4; UNIPARC:UPI0000051B81; GB:AE005672;

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP2216

Query Match

Best Local Similarity 76.7%; Score 89; DB 2; Length 392;

Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAAG 18

Db 304 GDYWGNGAQWATSAAG 321

|:||||| ||| |||||

|:||||| ||| |||||

RESULT 2

B98124

general stress protein GSP-781 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004

C:Accession: B98124

R:Hoekings, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bargett, S.; DeHoff, B.S.;

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.

A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: B98124

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <KUR>

A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMV4; UNIPARC:UPI0000051B81; GB:AE007317;

C:Genetics:

A:Gene: gsp-781

```
Query Match          76.7%;   Score 89;   DB 2;   Length 392;
Best Local Similarity 83.3%;   Pred. No. 0.0002; 2;   Indels 0;   Gaps 0;
Matches 15;   Conservative 1;   Mismatches 2;

QY 1 GNYWNGGGQWAASAAAAG 18
   |:|||||:|||||
Db 304 GDYWGNGAQWATSAAAAG 321

RESULT 3
C90029
hypothetical protein SA2097 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: C90029
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: C90029
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-166 <KUR>
A:Cross-references: UNIPROT:Q99RW9; UNIPARC:UPI000005434E; GB:BA000018; PID:g13702104; F
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2097

Query Match          52.6%;   Score 61;   DB 2;   Length 166;
Best Local Similarity 55.6%;   Pred. No. 0.44; 2;   Indels 0;   Gaps 0;
Matches 10;   Conservative 2;   Mismatches 6;

QY 1 GNYWNGGGQWAASAAAAG 18
   |:|||||:|||||
Db 82 GSTWGNWNAWAAQAAG 99

RESULT 4
F90028
hypothetical protein saa [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: F90028
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90028
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <KUR>
A:Cross-references: UNIPROT:Q99RX4; UNIPARC:UPI000005225C; GB:BA000018; PID:g13702099; F
A:Experimental source: strain N315
C:Genetics:
A:Gene: saa

Query Match          50.9%;   Score 59;   DB 2;   Length 267;
Best Local Similarity 55.6%;   Pred. No. 1.3; 2;   Indels 0;   Gaps 0;
Matches 10;   Conservative 2;   Mismatches 6;

QY 1 GNYWNGGGQWAASAAAAG 18
   |:|||||:|||||
Db 184 GSTWGNWNAWAAARAG 201

RESULT 5
G90061
hypothetical protein SA2353 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
```

```
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: G90061
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: G90061
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <KUR>
A:Cross-references: UNIPROT:Q53587; UNIPARC:UPI00000D7806; GB:BA000018; PID:g13702516; F
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2353

Query Match          50.0%;   Score 58;   DB 2;   Length 255;
Best Local Similarity 50.0%;   Pred. No. 1.6; 4;   Indels 0;   Gaps 0;
Matches 9;   Conservative 4;   Mismatches 5;

QY 1 GNYWNGGGQWAASAAAAG 18
   |:|||||:|||||
Db 172 GSTWGNWNAWAAASSG 189

RESULT 6
H36891
transfer complex protein trsg - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 09-Jul-2004
C:Accession: H36891
R:Morton, T.M.; Eaton, D.M.; Johnston, J.L.; Archer, G.L.
J. Bacteriol. 175, 4436-4447, 1993
A:Title: DNA sequence and units of transcription of the conjugative transfer gene comple
A:Reference number: A36891; MUID:93322322; PMID:7687249
A:Accession: H36891
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-358 <MOR>
A:Cross-references: UNIPROT:Q07727; UNIPARC:UPI00000B98DF; GB:L11998; NID:g310606; PIDN
C:Genetics:
A:Gene: trsg
C:Superfamily: Staphylococcus aureus transfer complex protein trsg

Query Match          49.1%;   Score 57;   DB 2;   Length 358;
Best Local Similarity 60.0%;   Pred. No. 3; 9;   Indels 0;   Gaps 0;
Matches 9;   Conservative 1;   Mismatches 5;

QY 4 WNGNGGQWAASAAAAG 18
   |:|||||:|||||
Db 271 WNGNGDWDGDNKAQAG 285

RESULT 7
C86863
N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) [imported] - Lactococcus lactis subsp.
N:Alternate names: N-acetylmuramidase
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: C86863
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86863
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-475 <STO>
A:Cross-references: UNIPROT:Q9CED5; UNIPARC:UPI00000C6B58; GB:AE005176; PID:g12724943; F
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: acmb
```


C;Keywords: hydrolase

Query Match 49.1%; Score 57; DB 2; Length 475;
Best Local Similarity 45.0%; Pred. No. 3.9;
Matches 9; Conservative 4; Mismatches 7; Indels

QY 1 GNYWNGGGQWAAASAAAAGRY 20
| : ||||| : | : | :
Db 385 GTHMNGGGEWGINAAQAQGYF 404

RESULT 8
T49625
glucan 1,4-alpha-glucosidase [imported] - *Neurospora crassa*
N;Alternate names: protein B5022.70
C;Species: *Neurospora crassa*
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2003
C;Accession: T49625
R;Schulte, U.; Algn. V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, A.
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49625
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-626 <SCH>
A;Cross-references: UNIPARC:UPI000006A8CB; EMBL:ALJ355932; GSPDB:GN00116; NCSP:B5022.70
A;Experimental source: BAC clone B5022; strain OR74A
C;Genetics:
A;Gene: NCSP:B5022.70
A;Map position: 6 ↓
A;Introns: 82/1
C;Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-

Query Match	47.0%	Score 54.5;	DB 2;	Length 636;
Best Local Similarity	57.9%;	Pred. No. 11;		
Match	11;	Conservative	2;	Mismatches
			5;	Indels
				Gaps
				1;

QY 3 YWG-NGGWAASAAAAAGRY 20
: : : : :
Db 323 FWGVNSGRTAGKAAAVGRY 34

RESULT 9
S36364
glucan 1,4-alpha-glucosidase (SC 3.2.1.1.3) precursor - *Neurospora crassa*
N;Alternate names: glucoamylase; glycoamylase
C;Species: *Neurospora crassa*
C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 02-Jun-2003
C;Accession: S36364; S13710; S13711; S25539
R;Stone, P.J.; Makoff, A.J.; Parish, J.H.; Radford, A.
Curr. Genet. 24, 205-211, 1993
A;Title: Cloning and sequence analysis of the glucoamylase gene of *Neurospora crassa*.
A;Reference number: S36364; MUID:94037144; PMID:8221928
A;Accession: S36364
A;Molecule type: DNA
A;Residues: 1-626 <STO>
A;Cross-references: UNIPARC:UPI000014A557; EMBL:X67291
R;Kroh-Luaz, S.I.; Parish, J.H.; Bleasaby, A.J.; Pappin, D.J.C.; Ainley, K.; Johansen, F. H.
Enzyme Microb. Technol. 11, 692-695, 1989
A;Title: Exported proteins of *Neurospora crassa*: l-glucoamylase.

C;Genetics:

A:Gene: gla-1
A:Introns: 82/2
C:Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; 9
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-35/Domain: propeptide #status predicted <PRO>
F:33-458/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
F:36-626/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>

Query Match 47.0%; Score 54.5; DB 2; Length 626;
Best Local Similarity 57.9%; Pred. No. 11;
Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

Qy 3 YWG-NGGWAASAAAAGRY 20
: : : : :
pb 323 FWGVNSGRTAGKAAAVGRY 341

RESULT 10
E95843
hypothetical protein SMD20010 [imported] - Sinorhizobium meliloti (strain 1021) megaplas
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C:Accession: E95843
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing end
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95843
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <KUR>
A:Cross-references: UNIPROT:Q9X2E9; UNIPARC:UPI00000CB3A1; GB:AL591985; PIDN:CAC48413.1
R:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:213668234; PMID:11474104

```
Query Match          46.6%; Score 54; DB 2; Length 355;
Best Local Similarity 69.2%; Pred. No. 7.4;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

QY 4 WNGGGQWAAASAAA 16
151 WATGGOWAAAFAA 163

RESULT 11

JN0097
secreted 45k protein precursor - Lactococcus lactis
C:Species: Lactococcus lactis
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: JN0097
R:van Asseldonk, M.; Rutten, G.; Oteman, M.; Siezen, R.J.; de Vos, W.M.; Simons, G.
Gene 95, 155-160, 1990
A:Title: Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis s
A:Reference number: JN0097; MUID:91071599; PMID:2123812
A:Accession: JN0097
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <VAN>
A:Cross-references: UNIPARC:UPI000016D72B; GB:W35374

Query Match 46.6%; Score 54; DB 2; Length 461;
Best Local Similarity 64.3%; Pred. No. 9.5;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GNGGQWAAASAAAAG 18
|||||:|
Db 376 GNGGQWASNGPAQG 389

RESULT 12

E85354
PR-1-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: E85354
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: E85354
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-161 <STO>
A:Cross-references: UNIPROT:Q9M0C8; UNIPARC:UPI00000A03A9; GB:NC_001268; NID:g7269932; E85354
A:Gene: At4g30320
A:Map position: 4
C:Superfamily: pathogenesis-related leaf protein

Query Match 45.7%; Score 53; DB 2; Length 161;
Best Local Similarity 50.0%; Pred. No. 4.8;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 YMGNGQWAAASAAAAG 18
:|:|:|:|:|
Db 81 FWGSGNRWGPSQRAYG 96

RESULT 13

E86903
hypothetical protein usp45 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: E86903
R:Solotkin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: E86903
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-456 <STO>
A:Cross-references: UNIPROT:Q9CDV1; UNIPARC:UPI00000C6BFC; GB:AE005176; PID:gl2725296; E86903
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: usp45

Query Match 45.7%; Score 53; DB 2; Length 456;
Best Local Similarity 64.3%; Pred. No. 13;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GNGGQWAAASAAAAG 18
|||||:|
Db 371 GNGGQWATNGPAQG 384

RESULT 14

S45608
light-harvesting protein II B-800/850 alpha chain - Rhodocyclus gelatinosus
N:Alternate names: peripheral antenna complex B800-850 alpha chain
C:Species: Rhodocyclus gelatinosus
C:Date: 10-Dec-1994 #sequence_revision 19-Jul-1996 #text_change 09-Jul-2004
C:Accession: S45608, S66177

R:Brumisholz, R.A.; Suter, F.; Zuber, H.

Eur. J. Biochem. 222, 667-675, 1994

A:Title: Structural and spectral characterisation of the antenna complexes of Rhodocycl

A:Reference number: S45608; MUID:94291666; PMID:8020505

A:Accession: S45608

A:Molecule type: protein

A:Residues: 1-71 <BRU>

A:Cross-references: UNIPROT:P77799; UNIPARC:UPI00000B40B7

A:Experimental source: DSM 149

A:Accession: S66177

A:Molecule type: protein

A:Residues: 1-71 <BR2>

A:Cross-references: UNIPARC:UPI00000B40B7

A:Experimental source: DSM 151

C:Superfamily: light-harvesting protein alpha chain

C:Keywords: antenna complex; bacteriochlorophyll; light-harvesting polypeptide; membran

Query Match 45.3%; Score 52.5; DB 2; Length 71;

Best Local Similarity 70.6%; Pred. No. 2.6;

Matches 12; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 GNYWNGGQWAAASAAA 17
| | | | | : | | | | |
Db 42 GAYW-NGGKKAASAAA 57

RESULT 15

D64754

probable oxidoreductase (EC 1.-.-.-) yagR [similarity] - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004

C:Accession: D64754

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: D64754

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-732 <BLAT>

A:Cross-references: UNIPROT:P77489; UNIPARC:UPI000013A0B1; GB:AE000136; NID:4

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yagR

C:Superfamily: carbon monoxide dehydrogenase molybdoprotein

C:Keywords: oxidoreductase

Query Match 44.8%; Score 52; DB 2; Length 732;

Best Local Similarity 50.0%; Pred. No. 27;

Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 GNGGQWAAASAAAAGRY 20
|:|:|:|:|:|:|:|:|:|
Db 505 GSGGQWGTSTSGVY 520

Search completed: February 11, 2006, 19:42:58

Job time : 1.41281 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 18:57:35 ; Search time 2.69869 Seconds
(without alignments)
5228.676 Million cell updates/sec
Title: US-10-797-821-41
Perfect score: 116
Sequence: 1 GNYWNGGQWAAASAAAAGRY 20
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	89.7	431	2	Q938V0_STRMU
2	104	89.7	431	2	Q938V3_STRMU
3	104	89.7	431	2	Q9AG98_STRMU
4	104	89.7	431	2	Q8DMW3_STRMU
5	104	89.7	432	2	Q938V1_STRMU
6	104	89.7	432	2	Q938V2_STRMU
7	91	78.4	169	2	Q8E3F4_STRAJ
8	90	77.6	486	2	Q8G4C1_BIFLO
9	89	76.7	392	2	Q8DMY4_STRR6
10	89	76.7	392	2	Q97N55_STRPN
11	85	73.3	169	2	Q8DXT4_STRAS
12	84	72.4	125	2	Q8RQM2_STRRT
13	84	72.4	398	2	Q5XEL1_STRP6
14	84	72.4	398	2	Q9A1Z8_STRP
15	84	72.4	398	2	Q7CNQ7_STRP8
16	84	72.4	398	2	Q8P318_STRP3
17	83	71.6	447	2	Q9AKA4_STRAG
18	83	71.6	447	2	Q8E2H1_STRAS
19	83	71.6	447	2	Q8E7X9_STRAJ
20	80	69.0	132	2	Q8RQM3_STRRE
21	78	67.2	125	2	Q8RQM1_STRMU
22	78	67.2	211	2	Q8DVU8_STRMU
23	76	65.5	128	2	Q8RQM0_STRRE
24	75	64.7	461	2	Q56SA7_STRTR
25	67	57.8	263	2	Q54487_STRAC
26	66	56.9	129	2	Q5M137_STRT1
27	66	56.9	474	2	Q5M6K4_STRT2
28	66	56.9	482	2	Q5M5M6_STRT2
29	66	56.9	485	2	Q5M212_STRT1
30	65	56.0	286	2	Q4L410_STRAHJ
31	65	56.0	544	2	Q840V8_STRMU

32	65	56.0	544	2	Q840W6_STRMU	Q840W6 streptococc
33	65	56.0	544	2	Q840X3_STRMU	Q840X3 streptococc
34	64.5	55.6	318	2	Q8G3U4_BIFLO	Q8G3U4 bifidobacte
35	64	55.2	215	2	Q4L8G1_STRAHJ	Q4L8G1 staphylococ
36	64	55.2	249	2	Q4L8F4_STRAHJ	Q4L8F4 staphylococ
37	62	53.4	59	2	Q5LY27_STRT1	Q5LY27 streptococc
38	62	53.4	59	2	Q5M2N2_STRT2	Q5M2N2 streptococc
39	61	52.6	157	2	Q5HLU8_STRABQ	Q5HLU8 staphylococ
40	61	52.6	157	2	Q8CMK5_STRABP	Q8CMK5 staphylococ
41	61	52.6	163	2	Q6GED0_STRAR	Q6GED0 staphylococ
42	61	52.6	166	2	Q6G718_STRAS	Q6G718 staphylococ
43	61	52.6	166	2	Q5HDQ5_STRAC	Q5HDQ5 staphylococ
44	61	52.6	166	2	Q7A060_STRAAW	Q7A060 staphylococ
45	61	52.6	166	2	Q7A2K7_STRAAW	Q7A2K7 staphylococ

ALIGNMENTS

RESULT 1
Q938V0_STRMU
ID Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5SM3;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR EMBL; AY046414; AAK94504.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS05911; CHAP; 1.
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8D8BC4609F CRC64;

Query Match 89.7%; Score 104; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAAG 18
|||||
Db 349 GNYWNGGQWAAASAAAAG 366

RESULT 2
Q938V3_STRMU
ID Q938V3_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V3_STRMU PRELIMINARY; PRT; 431 AA.
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Glucan-binding protein B.

OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;

[1]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=STJ32;

RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;

RT "Cloning of the gbpB gene from Streptococcus mutans.";

RL J. Dent. Res. 79:224-224(2000).

[2]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=STJ32;

RX MEDLINE=21481971; PubMed=11598068;

RY DOI=10.1128/IAI.69.11.6931-6941.2001;

RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,

RC Duncan M.J.;

RT "Cloning of the Streptococcus mutans gene encoding glucan binding

protein B and analysis of genetic diversity and protein production in

clinical isolates.";

RL Infect. Immun. 69:6931-6941(2001).

DR EMBL; AY046410; AAK94500.1; -; Genomic_DNA.

DR InterPro; IPR007921; CHAP.

DR InterPro; IPR009148; Siba.

DR Pfam; PF05257; CHAP; 1.

DR PRINTS; PR01852; SIBAPROTEIN.

DR PROSITE; PS50911; CHAP; 1.

SQ SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7B51 CRC64;

Query Match 89.7%; Score 104; DB 2; Length 431;

Best Local Similarity 100.0%; Pred. No. 1.6e-05;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18

|||||

Db 349 GNYWNGGQWAASAAAAG 366

RESULT 3

Q9AG98 STRMU

ID Q9AG98 STRMU PRELIMINARY; PRT; 431 AA.

AC Q9AG98;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)

DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).

GN Name=saga;

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

[1]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=GS-5;

RX MEDLINE=21153617; PubMed=11254612;

RY DOI=10.1128/IAI.69.4.2493-2501.2001;

RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;

RT "Identification of stress-responsive genes in Streptococcus mutans by

differential display reverse transcription-PCR.";

RL Infect. Immun. 69:2493-2501(2001).

[2]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=GS-5;

RX MEDLINE=21481977; PubMed=11598074;

RY DOI=10.1128/IAI.69.11.6987-6998.2001;

RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;

RT "A 60-kilodalton immunodominant glycoprotein is essential for cell

wall integrity and the maintenance of cell shape in Streptococcus

mutans.";

RL Infect. Immun. 69:6987-6998(2001).

[3]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=GS-5;

RA Chia J.-S., Chang L.-Y., Lee Y.-Y., Chen J.-Y.;

RL Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.

[4]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=3VF4;

RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;

RT "Cloning of the gbpB gene from Streptococcus mutans.";

RL J. Dent. Res. 79:224-224(2000).

[5]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=3VF4;

RX MEDLINE=21481971; PubMed=11598068;

RY DOI=10.1128/IAI.69.11.6931-6941.2001;

RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,

RC Duncan M.J.;

RT "Cloning of the Streptococcus mutans gene encoding glucan binding

protein B and analysis of genetic diversity and protein production in

clinical isolates.";

RL Infect. Immun. 69:6931-6941(2001).

DR EMBL; AF338445; AAK08104.1; -; Genomic_DNA.

DR EMBL; AY046411; AAK94501.1; -; Genomic_DNA.

DR InterPro; IPR007921; CHAP.

DR InterPro; IPR009148; Siba.

DR Pfam; PF05257; CHAP; 1.

DR PRINTS; PR01852; SIBAPROTEIN.

DR PROSITE; PS50911; CHAP; 1.

SQ SEQUENCE 431 AA; 44592 MW; 3EBE21FC5E47232E CRC64;

Query Match 89.7%; Score 104; DB 2; Length 431;

Best Local Similarity 100.0%; Pred. No. 1.6e-05;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18

|||||

Db 349 GNYWNGGQWAASAAAAG 366

RESULT 4

Q8DWM3 STRMU

ID Q8DWM3 STRMU PRELIMINARY; PRT; 431 AA.

AC Q8DWM3;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Putative secreted antigen GbpB/Saga; putative peptidoglycan

hydrolyase.

GN Name=gbpB; OrderedLocusNames=SMU.22;

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

[1]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=UA159 / ATCC 700610 / Serotype c;

RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;

RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.S.P.,

RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,

RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,

RA Ferretti J.J.;

RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental

pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

DR EMBL; AE014855; AAN57811.1; -; Genomic_DNA.

DR InterPro; IPR007921; CHAP.

DR InterPro; IPR009148; Siba.

DR Pfam; PF05257; CHAP; 1.

DR PRINTS; PR01852; SIBAPROTEIN.

DR PROSITE; PS50911; CHAP; 1.

DR Complete proteome.

SQ SEQUENCE 431 AA; 44620 MW; 2D1CA695248CCD3E CRC64;

Query Match 89.7%; Score 104; DB 2; Length 431;

Best Local Similarity 100.0%; Pred. No. 1.6e-05;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
|||||
Db 349 GNYWNGGQWAASAAAAG 366

RESULT 5

Q938V1_STRMU
ID Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F80ECB9A1F3BE4F CRC64;

Query Match 89.7%; Score 104; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
|||||
Db 350 GNYWNGGQWAASAAAAG 367

RESULT 6

Q938V2_STRMU
ID Q938V2_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046412; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AEE50E9 CRC64;

Query Match 89.7%; Score 104; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
|||||
Db 350 GNYWNGGQWAASAAAAG 367

RESULT 7

Q8E3F4_STRA3
ID Q8E3F4_STRA3 PRELIMINARY; PRT; 169 AA.
AC Q8E3F4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein gba1805.
GN OrderedLocusNames=gba1805;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766853; CAD47464.1; -; Genomic_DNA.
DR SAGaList; gba1805; -;
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIDA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 169 AA; 18304 MW; 8A159F753D747869 CRC64;

Query Match 78.4%; Score 91; DB 2; Length 169;
Best Local Similarity 83.3%; Pred. No. 0.00034;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAASAAAAG 18
|||||
Db 79 GNYWNGGQWAASAAAAG 96

RESULT 8

Q8G4C1_BIFLO
ID Q8G4C1_BIFLO PRELIMINARY; PRT; 486 AA.
AC Q8G4C1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Possible frag-related protein.
GN OrderedLocusNames=BL1467;
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787; DOI=10.1073/pnas.212527599;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Fessi G., Zwahlen R.D., Desiere F., Bork P., Delley M.,
RA Primore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RL to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AB014295; AA025262.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 486 AA; 51046 MW; D62375650A47B052 CRC64;

Query Match 77.6%; Score 90; DB 2; Length 486;
Best Local Similarity 83.3%; Pred. No. 0.0012;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWASAAAAG 18
Db 230 GSYWNGGDWAASATAAG 247

RESULT 9
Q8DMY4_STR6
ID Q8DMY4_STR6 PRELIMINARY; PRT; 392 AA.
AC Q8DMY4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE General stress protein GSP-781.
GN Name=gsp-781; OrderedLocusNames=gsp2021;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAnnis S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
RL EMBL; AB008566; AAL00823.1; -; Genomic_DNA.
DR PIR; B98124; B98124.
DR PIR; G95258; G95258.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 392 AA; 41697 MW; B5A5A860FAEA77DD CRC64;

```

```

Query Match 76.7%; Score 89; DB 2; Length 392;
Best Local Similarity 83.3%; Pred. No. 0.0014;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWASAAAAG 18
Db 304 GDYWGNGAQWATSAAAAG 321

RESULT 10
Q97N55_STRPN
ID Q97N55_STRPN PRELIMINARY; PRT; 392 AA.
AC Q97N55;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Secreted 45 kd protein.
GN OrderedLocusNames=SP2216;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzaple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
DR EMBL; AB007509; AAK76264.1; -; Genomic_DNA.
DR PIR; B98124; B98124.
DR PIR; G95258; G95258.
DR TIGR; SP2216; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 392 AA; 41697 MW; B5A5A860FAEA77DD CRC64;

Query Match 76.7%; Score 89; DB 2; Length 392;
Best Local Similarity 83.3%; Pred. No. 0.0014;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWASAAAAG 18
Db 304 GDYWGNGAQWATSAAAAG 321

RESULT 11
Q8DXT4_STRAS
ID Q8DXT4_STRAS PRELIMINARY; PRT; 169 AA.
AC Q8DXT4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein SAG1762.
GN OrderedLocusNames=SAG1762;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```



```

RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettein H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.P., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carthy H.A., Cline R.T., Van Aken S.E., Gill J., Scarcelli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014271; AN00625.1; -; Genomic_DNA.
DR TIGR; SAG1762; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 169 AA; 18390 MW; 8A1653A8B5B8B769 CRC64;

Query Match 73.3%; Score 85; DB 2; Length 169;
Best Local Similarity 82.4%; Pred. No. 0.0021;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAA 17
DB 79 GNYWNGGQWAAASAAAA 95

RESULT 12
OBQRM2_STRRT PRELIMINARY; PRT; 125 AA.
AC Q8RQM2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Streptococcus rattii.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1341;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=BHT;
RA Tamura H., Kato H.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083135; BAB88829.1; -; Genomic DNA.
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR002482; LysM.
DR Pfam; PF05257; CHAP; 1.
DR Pfam; PF01476; LysM; 1.
DR SMART; SM00257; LysM; 1.
DR PROSITE; PS0911; CHAP; 1.
KW Hypothetical protein.
FT NON TER
FT NON TER
SQ SEQUENCE 125 AA; 12798 MW; A26A0A2C34C1148A CRC64;

Query Match 72.4%; Score 84; DB 2; Length 125;
Best Local Similarity 77.8%; Pred. No. 0.0021;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAA 18
DB 89 GNYWNGGQWAAASAAAA 106

RESULT 13

```

```

Q5XEL1_STRP6 PRELIMINARY; PRT; 398 AA.
AC Q5XEL1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Secreted protein.
GN OrderedLocusNames=M6_Spy0017;
OS Streptococcus pyogenes (serotype M6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301450;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=MGAS10394;
RX PubMed=15272401; DOI=10.1086/422697;
RA Banks D.J., Porcella S.F., Barbian K.D., Beres S.B., Phillips L.E.,
RA Voyich J.M., Deleo F.R., Martin J.M., Somerville G.A., Musser J.M.;
RT "Progress toward characterization of the group A Streptococcus
RT metagenome: complete genome sequence of a macrolide-resistant serotype
RT M6 strain.";
RL J. Infect. Dis. 190:727-738(2004).
DR EMBL; CP000003; AAT86152.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 72.4%; Score 84; DB 2; Length 398;
Best Local Similarity 83.3%; Pred. No. 0.0064;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GNYWNGGQWAAASAAAA 18
DB 315 GNYWNGGQWAAASAAAA 332

RESULT 14
Q9A1Z8_STRPY PRELIMINARY; PRT; 398 AA.
AC Q9A1Z8; Q7BH59;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Secreted protein Siba precursor.
GN OrderedLocusNames=SPY0019;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN NUCLEOTIDE SEQUENCE.
RP Fagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;
RT "Identification and characterization of a novel secreted protein from
RT group A streptococcus.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006474; AAK33158.1; -; Genomic_DNA.
DR EMBL; AF319999; AAL73135.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.

```

```
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
KW Complete proteome; Signal.
FT SIGNAL 1 23 Potential.
SQ SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;

Query Match 72.4%; Score 84; DB 2; Length 398;
Best Local Similarity 83.3%; Pred. No. 0.0064;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
DB 315 GNNWNGGOWAYSQAAG 332

RESULT 15
O7CNQ7_STRP8
ID O7CNQ7_STRP8 PRELIMINARY; PRT; 398 AA.
AC O7CNQ7;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Putative secreted protein.
GN Ordered locus names=spyM18_0020;
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301451;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AE009955; AL96849.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SibA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS09111; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCED478CB863B60 CRC64;

Query Match 72.4%; Score 84; DB 2; Length 398;
Best Local Similarity 83.3%; Pred. No. 0.0064;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GNYWNGGOWAASAAAAG 18
DB 315 GNNWNGGOWAYSQAAG 332
```

Search completed: February 11, 2006, 19:39:13
Job time : 4.69869 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 2.09028 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-42

Perfect score: 121

Sequence: 1 NNHVSIVEAWSNDTPYLHDD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*
9: Geneseq2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	100.0	21	9	Adx37285 Streptoco
2	121	100.0	1590	7	Adp93657 Streptoco
3	121	100.0	1590	9	Adx37280 Streptoco
4	121	100.0	1592	2	Aar32925 Glucosyl
5	115	95.0	21	9	Adv68564 S. sobrin
6	115	95.0	21	9	Adv68562 S. downsi
7	110	90.9	1017	5	Aau79285 Streptoco
8	110	90.9	1475	5	Aau98031 S. mutans
9	110	90.9	1475	5	Aau98036 S. mutans
10	110	90.9	1475	5	Aau98037 S. mutans
11	110	90.9	1475	5	Aau98040 S. mutans
12	110	90.9	1475	5	Aau98035 S. mutans
13	110	90.9	1475	5	Aau98033 S. mutans
14	110	90.9	1475	5	Aau98034 S. mutans
15	110	90.9	1475	5	Aau98030 S. mutans
16	110	90.9	1475	5	Aau98039 S. mutans
17	110	90.9	1475	5	Aau98027 S. mutans
18	110	90.9	1475	5	Aau98032 S. mutans
19	110	90.9	1475	5	Aau98038 S. mutans
20	110	90.9	1475	7	Adp93654 Streptoco
21	110	90.9	1475	9	Adx37277 Streptoco
22	110	90.9	1476	5	Aau79284 Streptoco
23	104	86.0	21	9	Adv68551 S. mutans
24	104	86.0	21	9	Adv68560 S. mutans

25	101	83.5	1375	5	Aau98028 S. mutans
26	101	83.5	1375	5	Aau79288 Streptoco
27	101	83.5	1375	7	Adp93655 Streptoco
28	101	83.5	1375	9	Adx37278 Streptoco
29	76	62.8	1499	7	Adc54806 Protein S
30	73	60.3	21	9	Adv68561 S. mutans
31	73	60.3	1430	5	Aau98043 S. mutans
32	73	60.3	1430	5	Aau98044 S. mutans
33	73	60.3	1430	5	Aau98045 S. mutans
34	73	60.3	1430	5	Aau98042 S. mutans
35	73	60.3	1430	5	Aau98029 S. mutans
36	73	60.3	1430	5	Aau98041 S. mutans
37	73	60.3	1430	7	Adp93656 Streptoco
38	73	60.3	1430	9	Adx37279 Streptoco
39	68	56.2	223	6	Abb63230 Glucansuc
40	68	56.2	2835	5	Abb98574 Dextran s
41	68	56.2	2835	5	Abb55594 Amino aci
42	66	54.5	12	5	Abb98642 Dextrane-
43	66	54.5	1365	7	Adp93659 Streptoco
44	66	54.5	1365	9	Adx37282 Streptoco
45	64	52.9	223	6	Abb63228 Glucansuc

ALIGNMENTS

RESULT 1
ADX37285
ID ADX37285 standard; peptide; 21 AA.
XX
AC ADX37285;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus sobrinus glucan binding protein B peptide #1.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus sobrinus.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
(SMIT/) SMITH D J.
(TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
WPI; 2005-151644/16.
XX
New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC Class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.
XX
Example 5; SEQ ID NO 42; 73pp; English.
XX
The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a

```

CC Streptococcus sobrinus GbpB-derived peptide of the invention.
XX
SQ Sequence 21 AA;

  Query Match      100.0%; Score 121; DB 9; Length 21;
  Best Local Similarity 100.0%; Pred. No. 7, 9e-11;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPLYLHDD 21
Db 1 NNHVSIVEAWSNDTPLYLHDD 21

RESULT 2
ADD93657
ID ADD93657 standard; protein; 1590 AA.
XX
AC ADD93657;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus sobrinus glucosyltransferase-I.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.
XX
PN WQ2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
XX
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
PI WPI; 2003-845091/78.
XX
DR
XX
CC Composition useful as vaccines for dental caries comprises a fragment of
CC a glucan binding protein-B binding to a major histocompatibility complex
CC class II protein.
XX
PS Claim 16; Page 14; 49pp; English.
XX
CC The present sequence is the protein sequence of Streptococcus sobrinus
CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutants glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Diepitopic or
CC multiepitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
SQ Sequence 1590 AA;

  Query Match      100.0%; Score 121; DB 7; Length 1590;
  Best Local Similarity 100.0%; Pred. No. 1,1e-08;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPLYLHDD 21
Db 478 NNHVSIVEAWSNDTPLYLHDD 498

RESULT 3

```

```

ADX37280
ID ADX37280 standard; protein; 1590 AA.
XX
AC ADX37280;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus sobrinus glucan binding protein B.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus sobrinus.
XX
FN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
(SMIT/) SMITH D J.
(PAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
PI WPI; 2005-151644/16.
XX
DR
XX
CC New composition comprising a fragment of a glucan binding protein-B
CC (GbpB) that binds to MHC class II protein, and a biocompatible
CC microparticle, useful for producing an antibody (claimed) for immunizing
CC mammals against dental caries.
XX
PS Claim 7; SEQ ID NO 37; 73pp; English.
XX
CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus sobrinus GbpB protein of the invention.
XX
SQ Sequence 1590 AA;

  Query Match      100.0%; Score 121; DB 9; Length 1590;
  Best Local Similarity 100.0%; Pred. No. 1,1e-08;
  Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPLYLHDD 21
Db 478 NNHVSIVEAWSNDTPLYLHDD 498

RESULT 4
AAR32925
ID AAR32925 standard; protein; 1592 AA.
XX
AC AAR32925;
XX
DT 28-JUN-1993 (first entry)
XX
DE Glucosyltransferase I.
XX
KW GT-1; Streptococcus; dental; caries.
XX
OS Streptococcus sobrinus.
XX

```

PN JP05023188-A.
 XX
 PD 02-FEB-1993.
 XX
 PF 25-JUL-1991; | 91JP-00186592.
 XX
 PR 25-JUL-1991; 91JP-00186592.
 XX
 XX (KATO/) KATO K.
 PA (FUKU/) FUKUI I.
 XX
 XX WPI; 1993-079449/10.
 DR N-PSDB; AAQ37760.
 XX
 XX DNA sequence glucosyltransferase-I - comprises Streptococcus sobrinus
 PT DNA sequence with at least one nucleotide added or deleted.
 PT
 XX Claim 13; Page 15; 29pp; Japanese.
 XX
 XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating S.
 CC sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially
 CC digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp
 CC fragment was ligated into pUC18 and E. coli JM109 transformed with it. A
 CC GT-1 expressing clone was isolated and sequenced. The clone may be used
 CC in the development of a drug for dental caries
 XX
 SQ Sequence 1592 AA;
 Query Match 100.0%; Score 121; DB 2; Length 1592;
 Best Local Similarity 100.0%; Pred. No. 1.1e-08;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NNHVSIVEAWSNDTPYLHDD 21
 |||||
 DB 478 NNHVSIVEAWSNDTPYLHDD 498
 RESULT 5
 ADV68564
 ID ADV68564 standard; peptide; 21 AA.
 XX
 AC ADV68564;
 XX
 DT 24-FEB-2005 (first entry)
 XX
 DE S. sobrinus GFP2 EAW peptide.
 XX
 KW Immunogenicity; vaccine; glucosyltransferase; GFP; tooth disease;
 KW mouth disease; antibacterial; EAW peptide.
 XX
 OS Streptococcus sobrinus.
 XX
 PN US6827936-B1.
 XX
 PD 07-DEC-2004.
 XX
 PF 12-APR-1999; 99US-00290049.
 XX
 PR 13-APR-1998; 98US-0081550P.
 PR 08-JAN-1999; 99US-0115142P.
 XX
 XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
 PA Smith DJ, Taubman MA;
 PI WPI; 2005-009976/01.
 XX
 XX New immunogenic composition comprises a peptide corresponding to a
 PT subunit of glucosyltransferase, useful as glucosyltransferase subunit
 PT vaccine for preventing dental caries.
 XX
 PS Claim 2; SEQ ID NO 14; 18pp; English.
 XX

XX The present invention relates to a novel immunogenic composition which
 CC comprises a peptide corresponding to a subunit of glucosyltransferase
 CC (GTF) enzyme. The immunogenic composition is useful as a GTF subunit
 CC vaccine for preventing dental caries. The present sequence is the
 CC Streptococcus sobrinus GFP2 EAW [(beta, alpha)8 barrel domain beta 5
 CC strand associated catalytic region] peptide.
 XX
 SQ Sequence 21 AA;
 Query Match 95.0%; Score 115; DB 9; Length 21;
 Best Local Similarity 100.0%; Pred. No. 6.1e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NNHVSIVEAWSNDTPYLHD 20
 |||||
 DB 2 NNHVSIVEAWSNDTPYLHD 21
 RESULT 6
 ADV68562
 ID ADV68562 standard; peptide; 21 AA.
 XX
 AC ADV68562;
 XX
 DT 24-FEB-2005 (first entry)
 XX
 DE S. downei GFP-I EAW peptide.
 XX
 KW Immunogenicity; vaccine; glucosyltransferase; GFP; tooth disease;
 KW mouth disease; antibacterial; EAW peptide.
 XX
 OS Streptococcus downei.
 XX
 PN US6827936-B1.
 XX
 PD 07-DEC-2004.
 XX
 PF 12-APR-1999; 99US-00290049.
 XX
 PR 13-APR-1998; 98US-0081550P.
 PR 08-JAN-1999; 99US-0115142P.
 XX
 XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
 PA Smith DJ, Taubman MA;
 PI WPI; 2005-009976/01.
 XX
 XX New immunogenic composition comprises a peptide corresponding to a
 PT subunit of glucosyltransferase, useful as glucosyltransferase subunit
 PT vaccine for preventing dental caries.
 XX
 PS Claim 2; SEQ ID NO 12; 18pp; English.
 XX
 XX The present invention relates to a novel immunogenic composition which
 CC comprises a peptide corresponding to a subunit of glucosyltransferase
 CC (GTF) enzyme. The immunogenic composition is useful as a GTF subunit
 CC vaccine for preventing dental caries. The present sequence is the
 CC Streptococcus downei GFP-I EAW [(beta, alpha)8 barrel domain beta 5 strand
 CC associated catalytic region] peptide.
 XX
 SQ Sequence 21 AA;
 Query Match 95.0%; Score 115; DB 9; Length 21;
 Best Local Similarity 100.0%; Pred. No. 6.1e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NNHVSIVEAWSNDTPYLHD 20
 |||||
 DB 2 NNHVSIVEAWSNDTPYLHD 21

QY 1 NNHVSIVEAWSNDNTPYLHDD 21
 Db 482 NDHLSILEAWSNDNTPYLHDD 502

RESULT 9
 AAU98036
 ID AAU98036 standard; protein; 1475 AA.
 AC AAU98036;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant D567T/D571K.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
 FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"
 FT US2002031826-A1.
 PN 14-MAR-2002.
 PD 19-DEC-2000; 2000US-00740274.
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX Nichols SE;
 XX WPI; 2002-414332/44.
 XX
 FT Glucosyltransferase B or D protein useful for producing a glucan useful
 FT as substitutes for and additions to modified starch and latexes in paper
 FT manufacture, comprises mutations in specific positions.
 PT
 PT
 PS Claim 36; Page; 4app; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I48V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

comprising the glucan (paper sizing/coating agent). The vector is useful
 for producing a glucan in a plant. The method comprises transforming a
 plant cell with the vector, growing the plant cell under plant growing
 conditions to produce a regenerated plant and inducing expression of the
 polynucleotide for a time sufficient to produce the glucan in the
 regenerated plant, where the vector contains a transit sequence from
 ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 chlorophyll AB binding protein to produce a transgenic plant, and glucan
 is produced in the amyloplast of potato or the vacuole of sugar beet.
 Glucans are useful as substitutes for and additions to modified starch
 and latexes in paper manufacture. Unlike prior art techniques, which
 require input materials that produce chemical effluents, paper
 manufacture utilising the glucan produced by GTF, which utilises
 biologically produced input materials, is more cost-effective and
 environmentally friendly. Moreover, glucans also exhibit thermoplastic
 properties and impart gloss to the paper during coating step. The present
 sequence represents a GTFB mutant of the invention. Note: The present
 sequence is not shown in the specification but was created by the indexer
 using the GTFB sequence appearing as AAU98027 and the information in
 claim 36

XX
 XX Sequence 1475 AA;
 SQ

Query Match 90.9%; Score 110; DB 5; Length 1475;
 Best Local Similarity 85.7%; Pred. NO. 4.5e-07;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDNTPYLHDD 21
 Db 482 NDHLSILEAWSNDNTPYLHDD 502

RESULT 10
 AAU98037
 ID AAU98037 standard; protein; 1475 AA.
 AC AAU98037;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant D567T/D571K/K1014T.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
 FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"
 FT US2002031826-A1.
 PN 14-MAR-2002.
 PD 19-DEC-2000; 2000US-00740274.
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 16-JAN-1998; 98US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX Nichols SE;
 XX WPI; 2002-414332/44.
 XX
 FT Glucosyltransferase B or D protein useful for producing a glucan useful
 FT as substitutes for and additions to modified starch and latexes in paper
 FT manufacture, comprises mutations in specific positions.
 PT
 PT
 PS Claim 36; Page; 4app; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I48V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper

PI Nichols SE;
 XX WPI; 2002-414332/44.
 XX
 XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 XX Claim 36; Page; 44pp; English.
 PS
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 1448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K799Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 XX
 SQ Sequence 1475 AA;
 Query Match 90.9%; Score 110; DB 5; Length 1475;
 Best Local Similarity 85.7%; Pred. No. 4.5e-07;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NNHVSIVEAWSNDPTPLHDD 21
 Db 482 NDHLSILEAWSNDPTPLHDD 502
 RESULT 11
 AAU98040
 ID AAU98040 standard; protein; 1475 AA.
 XX
 XX AAU98040;
 AC
 XX 27-AUG-2002 (first entry)
 DT
 XX S. mutans glucosyltransferase GTFB mutant K779Q.
 DE
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XI Streptococcus mutans.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 779
 FT /note= "Wild-type Lys substituted by Gln"
 XX
 XX US2002031826-A1.
 XX
 XX 14-MAR-2002.
 XX
 XX 19-DEC-2000; 2000US-00740274.
 XX
 XX 07-JUN-1995; 95US-00478704.
 XX 07-JUN-1995; 95US-00482711.
 XX 16-JAN-1995; 95US-00485243.
 XX 16-JAN-1998; 98US-00007999.
 XX 16-JAN-1998; 98US-00008172.
 XX 20-JAN-1998; 98US-00009620.
 XX 11-DEC-1998; 98US-00210361.
 XX (NICH/) NICHOLS S E.
 XX
 XX Nichols SE;
 WPI; 2002-414332/44.
 Glucosyltransferase B or D protein useful for producing a glucan useful
 as substitutes for and additions to modified starch and latexes in paper
 manufacture, comprises mutations in specific positions.
 Claim 36; Page; 44pp; English.
 The invention an isolated protein comprising a glucosyltransferase (GTF)
 B polypeptide having changes at position from 1448V, D457N, D567T,
 K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 I448V/D457N/D567T/D571K/K799Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 complementary polynucleotide, a ribonucleic acid sequence encoding the
 GTF mutant, an expression cassette comprising the polynucleotide operably
 linked to a promoter, a vector comprising the expression cassette, host
 cell introduced with the vector, a transgenic plant comprising the
 vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 coating composition comprising a glucan produced in a plant transformed
 with a gene encoding the mutant GTF, wild type or, starch, a latex,
 thermoplastic molecule or their combinations or glucan and starch where
 the glucan is produced in the amyloplast and/or vacuole or a maize line
 deficient in starch biosynthesis, transformed with a gene encoding a
 glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 comprising the glucan (paper sizing/coating agent). The vector is useful
 for producing a glucan in a plant. The method comprises transforming a
 plant cell with the vector, growing the plant cell under plant growing
 conditions to produce a regenerated plant and inducing expression of the
 polynucleotide for a time sufficient to produce the glucan in the
 regenerated plant, where the vector contains a transit sequence from
 ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 chlorophyll AB binding protein to produce a transgenic plant, and glucan
 is produced in the amyloplast of potato or the vacuole of sugar beet.
 Glucans are useful as substitutes for and additions to modified starch
 and latexes in paper manufacture. Unlike prior art techniques, which
 require input materials that produce chemical effluents, paper
 manufacture utilising the glucan produced by GTF, which utilises
 biologically produced input materials, is more cost-effective and
 environmentally friendly. Moreover, glucans also exhibit thermoplastic
 properties and impart gloss to the paper during coating step. The present
 sequence represents a GTFB mutant of the invention. Note: The present
 sequence is not shown in the specification but was created by the indexer
 using the GTFB sequence appearing as AAU98027 and the information in
 claim 36
 Sequence 1475 AA;

SQ Sequence 1475 AA;

Query Match 90.9%; Score 110; DB 5; Length 1475;
 Best Local Similarity 85.7%; Pred. No. 4.5e-07;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDNDTPYLHDD 21
 Db 482 NDHLSILEAWSNDNDTPYLHDD 502

RESULT 12
 AAU98035
 ID AAU98035 standard; protein; 1475 AA.
 XX
 AC AAU98035;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant D457N/D571K.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"
 FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
 FT XX
 PN US2002031826-A1.
 XX
 PD 14-MAR-2002.
 XX
 PF 19-DEC-2000; 2000US-00740274.
 XX
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.
 XX
 PI Nichols SE;
 XX
 DR WPI; 2002-414332/44.
 XX
 PT Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 PS Claim 36; Page; 44pp; English.
 XX
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed

CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 CC
 XX
 SQ Sequence 1475 AA;
 Query Match 90.9%; Score 110; DB 5; Length 1475;
 Best Local Similarity 85.7%; Pred. No. 4.5e-07;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDNDTPYLHDD 21
 Db 482 NDHLSILEAWSNDNDTPYLHDD 502

RESULT 13
 AAU98033
 ID AAU98033 standard; protein; 1475 AA.
 XX
 AC AAU98033;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE S. mutans glucosyltransferase GTFB mutant K1014T.
 XX
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.
 XX
 OS Streptococcus mutans.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"
 FT XX
 PN US2002031826-A1.
 XX
 PD 14-MAR-2002.
 XX
 PF 19-DEC-2000; 2000US-00740274.
 XX
 PR 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.
 XX
 PA (NICH/) NICHOLS S E.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 0.45978 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-42

Perfect score: 121

Sequence: 1 NNHVSIVEAWSNDTPYLHDD 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115	95.0	21	2	US-09-290-049A-12
2	115	95.0	21	2	US-09-290-049A-14
3	110	90.9	1475	2	US-09-007-999-2
4	110	90.9	1475	2	US-09-210-361-2
5	110	90.9	1475	2	US-09-740-274-2
6	104	86.0	21	2	US-09-290-049A-1
7	104	86.0	21	2	US-09-290-049A-10
8	101	83.5	1375	2	US-09-210-361-4
9	101	83.5	1375	2	US-09-740-274-4
10	73	60.3	21	2	US-09-290-049A-11
11	73	60.3	545	2	US-09-604-957-4
12	73	60.3	545	2	US-09-995-749A-10
13	73	60.3	1430	2	US-09-008-172-2
14	73	60.3	1430	2	US-09-210-361-6
15	73	60.3	1430	2	US-09-740-274-6
16	61	50.4	21	2	US-09-290-049A-13
17	61	50.4	522	2	US-09-995-749A-11
18	61	50.4	523	2	US-09-604-957-5
19	57	47.1	1577	1	US-08-793-824-2
20	51	42.1	584	2	US-09-604-957-6
21	51	42.1	584	2	US-09-995-749A-12
22	51	42.1	2057	2	US-09-499-203-2
23	47	38.8	535	2	US-09-604-957-7
24	47	38.8	535	2	US-09-995-749A-13
25	47	38.8	749	2	US-09-562-737-96
26	47	38.8	1278	2	US-09-604-957-3
27	47	38.8	1781	2	US-09-995-749A-2

28	47	38.8	2465	1	US-08-596-291-3	Sequence 3, Appli
29	47	38.8	2465	2	US-09-100-804-3	Sequence 3, Appli
30	47	38.8	2466	2	US-09-080-855-12	Sequence 12, Appli
31	47	38.8	2466	2	US-09-566-076-12	Sequence 12, Appli
32	47	38.8	2466	4	PCT-US94-09943-2	Sequence 2, Appli
33	47	38.8	2485	2	US-09-290-640-46	Sequence 46, Appli
34	47	38.8	2485	2	US-09-665-615B-46	Sequence 46, Appli
35	46	38.0	261	2	US-09-902-540-13119	Sequence 13119, A
36	46	38.0	574	2	US-10-104-047-3209	Sequence 3209, Ap
37	46	38.0	700	2	US-08-931-952-2	Sequence 2, Appli
38	46	38.0	700	2	US-08-272-247-2	Sequence 2, Appli
39	46	38.0	700	4	PCT-US95-08560-2	Sequence 2, Appli
40	45.5	37.6	151	2	US-09-198-452A-815	Sequence 815, App
41	45.5	37.6	205	2	US-09-438-185A-767	Sequence 767, App
42	45	37.2	199	2	US-09-605-703B-2226	Sequence 2226, Ap
43	45	37.2	445	4	PCT-US94-05387-7	Sequence 7, Appli
44	45	37.2	895	1	US-08-123-161A-8	Sequence 8, Appli
45	45	37.2	895	1	US-08-483-278-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-09-290-049A-12
; Sequence 12, Application US/09290049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. downei
US-09-290-049A-12

Query Match 95.0%; Score 115; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.2e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHD 20
DB 2 NNHVSIVEAWSNDTPYLHD 21

RESULT 2

US-09-290-049A-14
; Sequence 14, Application US/09290049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CARRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19


```
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EAW peptide
US-09-290-049A-1

Query Match      86.0%; Score 104; DB 2; Length 21;
Best Local Similarity 85.0%; Pred. No. 1.1e-09;
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NNHVSIVEAWSNDNDTPYLHD 20
Db 2 NDHLSILEAWSNDNDTPYLHD 21

RESULT 7
US-09-290-049A-10
; Sequence 10, Application US/09290049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049A-10

Query Match      86.0%; Score 104; DB 2; Length 21;
Best Local Similarity 85.0%; Pred. No. 1.1e-09;
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NNHVSIVEAWSNDNDTPYLHD 20
Db 2 NDHLSILEAWSNDNDTPYLHD 21

US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match      83.5%; Score 101; DB 2; Length 1375;
Best Local Similarity 81.0%; Pred. No. 4e-07;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NNHVSIVEAWSNDNDTPYLHDD 21
Db 508 NDHLSILEAWSNDNDTPYLHDD 528

RESULT 9
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match      83.5%; Score 101; DB 2; Length 1375;
Best Local Similarity 81.0%; Pred. No. 4e-07;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NNHVSIVEAWSNDNDTPYLHDD 21
Db 508 NDHLSILEAWSNDNDTPYLHDD 528

RESULT 10
US-09-290-049A-11
; Sequence 11, Application US/09290049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
```

```
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049A-11

Query Match      60.3%; Score 73; DB 2; Length 21;
Best Local Similarity 68.4%; Pred. No. 7.9e-05;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  2 NHVSIVEAWSNDTPYLHD 20
Db   3 NHLSILEAWSNDPQYNKD 21

RESULT 11
US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-604-957-4

Query Match      60.3%; Score 73; DB 2; Length 545;
Best Local Similarity 68.4%; Pred. No. 0.0033;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  2 NHVSIVEAWSNDTPYLHD 20
Db   3 NHLSILEAWSNDPQYNKD 21

RESULT 12
US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-995-749A-10

Query Match      60.3%; Score 73; DB 2; Length 545;
Best Local Similarity 68.4%; Pred. No. 0.0033;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  2 NHVSIVEAWSNDTPYLHD 20
Db   77 NHLSILEAWSNDPQYNKD 95

RESULT 13
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match      60.3%; Score 73; DB 2; Length 1430;
Best Local Similarity 68.4%; Pred. No. 0.0099;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  2 NHVSIVEAWSNDTPYLHD 20
Db   497 NHLSILEAWSNDPQYNKD 515

RESULT 14
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match      60.3%; Score 73; DB 2; Length 1430;
Best Local Similarity 68.4%; Pred. No. 0.0099;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 NHVSIVEAWSNDTPYLHD 20
 ||:|||||
 Db 497 NHLSILEAWSNDPQYNKD 515

RESULT 15

US-09-740-274-6
 ; Sequence 6, Application US/09740274
 ; Patent No. 6465203
 ; GENERAL INFORMATION:
 ; APPLICANT: Nichols, Scott E.
 ; TITLE OF INVENTION: Glucan-containing Compositions and Paper
 ; FILE REFERENCE: 0357CRD
 ; CURRENT APPLICATION NUMBER: US/09740,274
 ; CURRENT FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: 09/210,361
 ; PRIOR FILING DATE: 1998-12-11
 ; PRIOR APPLICATION NUMBER: 09/007,999
 ; PRIOR FILING DATE: 1998-01-16
 ; PRIOR APPLICATION NUMBER: 08/478,704
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: 09/009,620
 ; PRIOR FILING DATE: 1998-01-20
 ; PRIOR APPLICATION NUMBER: 08/485,243
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: 09/008,172
 ; PRIOR FILING DATE: 1998-01-16
 ; PRIOR APPLICATION NUMBER: 08/482,711
 ; PRIOR FILING DATE: 1995-06-07
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 1430
 ; TYPE: PRT
 ; ORGANISM: streptococcus mutans
 US-09-740-274-6

Query Match 60.3%; Score 73; DB 2; Length 1430;
 Best Local Similarity 68.4%; Pred. No. 0.0099;
 Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 NHVSIVEAWSNDTPYLHD 20
 ||:|||||
 Db 497 NHLSILEAWSNDPQYNKD 515

Search completed: February 11, 2006, 19:46:19
 Job time : 0.45978 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 1.87963 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-42
Perfect score: 121
Sequence: 1 NNHVSIVEAWSNDTPYLHDD 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:*

- 1: /cgn2_6/prodata1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata1/pubpaa/US08_PUBCOMB.pep:*
- 3: /cgn2_6/prodata1/pubpaa/US09_PUBCOMB.pep:*
- 4: /cgn2_6/prodata1/pubpaa/US10_PUBCOMB.pep:*
- 5: /cgn2_6/prodata1/pubpaa/US10B_PUBCOMB.pep:*
- 6: /cgn2_6/prodata1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121	100.0	21	US-10-797-821-42	Sequence 42, Appl
2	121	100.0	1590	US-10-383-930-37	Sequence 37, Appl
3	121	100.0	1590	US-10-797-821-37	Sequence 37, Appl
4	115	95.0	21	US-10-797-302-12	Sequence 12, Appl
5	115	95.0	21	US-10-797-302-14	Sequence 14, Appl
6	110	90.9	1475	US-09-740-274-2	Sequence 2, Appl
7	110	90.9	1475	US-10-383-930-34	Sequence 34, Appl
8	110	90.9	1475	US-10-797-821-34	Sequence 34, Appl
9	104	86.0	21	US-10-797-302-1	Sequence 1, Appl
10	101	83.5	1375	US-09-740-274-4	Sequence 4, Appl
11	101	83.5	1375	US-10-383-930-35	Sequence 35, Appl
12	101	83.5	1375	US-10-797-821-35	Sequence 35, Appl
13	95	78.5	21	US-10-797-302-10	Sequence 10, Appl
14	73	60.3	21	US-10-797-302-11	Sequence 11, Appl
15	73	60.3	545	US-09-995-749A-10	Sequence 10, Appl
16	73	60.3	1430	US-09-740-274-6	Sequence 6, Appl
17	73	60.3	1430	US-10-383-930-36	Sequence 36, Appl
18	73	60.3	1430	US-10-797-821-36	Sequence 36, Appl
19	68	56.2	223	US-10-484-218-10	Sequence 10, Appl
20	66	54.5	1365	US-10-383-930-39	Sequence 39, Appl
21	66	54.5	1365	US-10-797-821-39	Sequence 39, Appl
22	64	52.9	223	US-10-484-218-6	Sequence 6, Appl
23	61	50.4	21	US-10-797-302-13	Sequence 13, Appl
24	61	50.4	522	US-09-995-749A-11	Sequence 11, Appl
25	61	50.4	1554	US-10-383-930-38	Sequence 38, Appl
26	61	50.4	1554	US-10-797-821-38	Sequence 38, Appl
27	60	49.6	1006	US-10-484-218-22	Sequence 22, Appl

28 58 47.9 224 5 US-10-484-218-4 Sequence 4, Appl
29 58 47.9 1497 5 US-10-484-218-18 Sequence 18, Appl
30 58 47.9 1595 5 US-10-484-218-20 Sequence 20, Appl
31 56.5 46.7 266 4 US-10-437-963-180798 Sequence 180798,
32 56.5 46.7 1398 4 US-10-437-963-180124 Sequence 180124,
33 56 46.3 1518 4 US-10-383-930-40 Sequence 40, Appl
34 56 46.3 1518 5 US-10-797-821-40 Sequence 40, Appl
35 52.5 43.4 1244 4 US-10-437-963-180121 Sequence 180121,
36 51 42.1 584 3 US-09-995-749A-12 Sequence 12, Appl
37 51 42.1 2057 4 US-10-417-280A-2 Sequence 2, Appl
38 50 41.3 521 4 US-10-282-122A-61723 Sequence 61723, A
39 49.5 40.9 351 4 US-10-112-944-246 Sequence 246, App
40 49.5 40.9 539 3 US-09-764-875-762 Sequence 762, App
41 49.5 40.9 847 4 US-10-094-749-3045 Sequence 3045, Ap
42 49.5 40.9 1604 3 US-09-888-615-73 Sequence 73, Appl
43 48.5 40.1 401 4 US-10-158-057-222 Sequence 22, App
44 47 38.8 121 4 US-10-424-599-150233 Sequence 150233,
45 47 38.8 221 5 US-10-484-218-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-10-797-821-42
; Sequence 42, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 42
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus

US-10-797-821-42

Query Match 100.0%; Score 121; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.8e-11;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
| | | | | | | | | | | | | | | | | | | | |
Db 1 NNHVSIVEAWSNDTPYLHDD 21

RESULT 2

US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930

; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match 100.0%; Score 121; DB 4; Length 1590;
Best Local Similarity 100.0%; Pred. No. 5.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NNHVSIVEAWSNDTPLYHDD 21
Db 478 NNHVSIVEAWSNDTPLYHDD 498
|||||

RESULT 3
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match 100.0%; Score 121; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 5.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NNHVSIVEAWSNDTPLYHDD 21
Db 478 NNHVSIVEAWSNDTPLYHDD 498
|||||

RESULT 4
US-10-797-302-12
; Sequence 12, Application US/10797302
; Publication No. US20050026271A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-019
; CURRENT APPLICATION NUMBER: US/10/797,302
; CURRENT FILING DATE: 2004-03-09

; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-797-302-12

Query Match 95.0%; Score 115; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NNHVSIVEAWSNDTPLYHDD 20
Db 2 NNHVSIVEAWSNDTPLYHDD 21
|||||

RESULT 5
US-10-797-302-14
; Sequence 14, Application US/10797302
; Publication No. US20050026271A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-019
; CURRENT APPLICATION NUMBER: US/10/797,302
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-302-14

Query Match 95.0%; Score 115; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NNHVSIVEAWSNDTPLYHDD 20
Db 2 NNHVSIVEAWSNDTPLYHDD 21
|||||

RESULT 6
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1

; ORGANISM: Streptococcus mutans
US-10-797-302-10

Query Match 78.5%; Score 95; DB 5; Length 21;
Best Local Similarity 80.0%; Pred. No. 3.8e-07;
Matches 16; Conservative 3; Mismatches 1; Indels

Qy	1	NNHVSIVEAWS	20
		NDNTPYLHD	
		: : : : : :	
Db	2	NDHLSILEAWS <td>21</td>	21
		NDNTPYLHD	

RESULT 14

```

US-10-797-302-11
; Sequence 11, Application US/10797302
; Publication No. US20050026271A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-019
; CURRENT APPLICATION NUMBER: US/10/797,302
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-302-11

```

Query Match	60.3%;	Score 73;	DB 5;	Length 21;
Best Local Similarity	68.4%;	Pred. No. 0.00075;		
Matches 13;	Conservative	2;	Mismatches 4;	Indels 0;
				Gaps 0;

Qy 2 NHVSIVEAWSNDTPYLHD 20
|||:|||||
Db 3 NHLSILEAWSNDPOYNKD 21

RESULT 15

```

US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. US2002015556A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBERT
; APPLICANT: RAHROUFI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 10
; LENGTH: 545

```

THIS PAGE BLANK (USPTO)


```
; APPLICANT: Tomoyasu Sugiyama
; APPLICANT: Ai Wakamatsu
; APPLICANT: Keiichi Nagai
; APPLICANT: Tetsuji Otsuki
; APPLICANT: Shin-Ichi Funahashi
; APPLICANT: Chiaki Senoo
; APPLICANT: Jun-Ichi Nezu
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEIN KINASE/PROTEIN
; TITLE OF INVENTION: PHOSPHATASE
; FILE REFERENCE: 06501-099002
; CURRENT APPLICATION NUMBER: US/11/109,156
; CURRENT FILING DATE: 2005-04-19
; PRIOR APPLICATION NUMBER: US/10/060,065
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: PCT/JP00/05061
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: US 60/183,322
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: JP 11-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1267
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-11-109-156-35

Query Match 38.8%; Score 47; DB 7; Length 1267;
Best Local Similarity 47.6%; Pred. No. 20;
Matches 10; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

QY 3 HVSIVE--AWSNDTPYLHDD 21
Db 1149 HSHLNFTAWPDHDTESQPDD 1169

RESULT 3
US-11-150-406-2
; Sequence 2, Application US/11/150406
; Publication No. US20050250164A1
; GENERAL INFORMATION:
; APPLICANT: Muschler, John L
; APPLICANT: Bissell, Mina J
; TITLE OF INVENTION: Design of Novel Assays Based on the Newly Found Role of
; TITLE OF INVENTION: Dystroglycan and Alpha-Dystroglycan Proteolysis in Tumor Cell
; TITLE OF INVENTION: Growth
; FILE REFERENCE: IB-1398A
; CURRENT APPLICATION NUMBER: US/11/150,406
; CURRENT FILING DATE: 2005-06-09
; PRIOR APPLICATION NUMBER: 60/151,766
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 09/652,493
; PRIOR FILING DATE: 2000-08-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 895
; TYPE: PRT
; ORGANISM: homo sapiens
; ORGANISM: homo sapiens
US-11-150-406-2

Query Match 37.2%; Score 45; DB 7; Length 895;
Best Local Similarity 35.7%; Pred. No. 29;
Matches 10; Conservative 4; Mismatches 6; Indels 8; Gaps 1;

; APPLICANT: 2 NHVSIVEAW-----SDNDTPYLHDD 21
; Db 501 NHIDRVDAWVGTYFEVKIPSDTFYDHD 528

RESULT 4
US-10-821-234-1129
; Sequence 1129, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1129
; LENGTH: 920
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-10-821-234-1129

Query Match 37.2%; Score 45; DB 6; Length 920;
Best Local Similarity 35.7%; Pred. No. 29;
Matches 10; Conservative 4; Mismatches 6; Indels 8; Gaps 1;

QY 2 NHVSIVEAW-----SDNDTPYLHDD 21
Db 526 NHIDRVDAWVGTYFEVKIPSDTFYDHD 553

RESULT 5
US-11-169-041-174
; Sequence 174, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 10001 NP
; CURRENT APPLICATION NUMBER: US/11/169,041
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 174
; LENGTH: 1366
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-11-169-041-174

Query Match 37.2%; Score 45; DB 7; Length 1366;
Best Local Similarity 47.4%; Pred. No. 44;
Matches 9; Conservative 1; Mismatches 7; Indels 2; Gaps 1;

QY 2 NHVSIVEAWSNDTPYLHD 20
Db 130 NOVSTI--WDDNPAFSTHD 146

RESULT 6
US-11-010-239-113
; Sequence 113, Application US/11010239
```

```

; Publication No. US20060015970A1
; GENERAL INFORMATION:
; APPLICANT: ROGER PENNELL
; APPLICANT: Jack OKAMURO
; APPLICANT: Richard SCHNEEBERGER
; APPLICANT: Yiwen FANG
; APPLICANT: Shing KWOK
; APPLICANT: Diane JOFUKU
; APPLICANT: Edward A. KIEGLE
; APPLICANT: Jonathan DONSON
; APPLICANT: Nestor APUYA
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND POLYPEPTIDES ENCODED THEREBY USEFUL FOR
; TITLE OF INVENTION: MODIFYING PLANT CHARACTERISTICS
; FILE REFERENCE: 2750-1585PUS2
; CURRENT FILING DATE: 2004-12-09
; PRIOR APPLICATION NUMBER: US 60/529,352
; PRIOR FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 113
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(435)
; OTHER INFORMATION: 12600234_protein_ID_12600235
US-11-010-239-113

Query Match      36.4%; Score 44; DB 7; Length 435;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      6 IVEAWSNDTPY 17
DB      344 VLEAWSDKTSPP 355

RESULT 7
US-11-112-882-29
; Sequence 29, Application US/11112882
; Publication No. US20050273885A1
; GENERAL INFORMATION:
; APPLICANT: Commonwealth Scientific and Industrial Research Organisation
; TITLE OF INVENTION: Synthesis of Long-Chain Polyunsaturated Fatty Acids in Recombinant
; FILE REFERENCE: 503244
; CURRENT FILING DATE: 2005-04-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-112-882-29

Query Match      36.0%; Score 43.5; DB 7; Length 443;
Best Local Similarity 45.5%; Pred. No. 23;
Matches 10; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY      1 NNHVSIVEAW-SDNDTPYLHDD 21
DB      398 NACVKYKWKCKENNLPLYLDD 419

RESULT 8
US-10-873-528-192
; Sequence 192, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG

```

```

; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 192
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-192

Query Match      35.5%; Score 43; DB 6; Length 318;
Best Local Similarity 40.9%; Pred. No. 20;
Matches 9; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY      2 NHVSIVEAWSNDTP--YLHDD 21
DB      170 NGLTETGQKNDTGYWVHSD 191

RESULT 9
US-11-052-554A-233
; Sequence 233, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 233
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-233

Query Match      35.5%; Score 43; DB 7; Length 318;
Best Local Similarity 40.9%; Pred. No. 20;
Matches 9; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

QY      2 NHVSIVEAWSNDTP--YLHDD 21
DB      170 NGLTETGQKNDTGYWVHSD 191

RESULT 10
US-10-995-561-890
; Sequence 890, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702

```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 890
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-890

Query Match          35.5%; Score 43; DB 6; Length 617;
Best Local Similarity 38.9%; Pred. No. 39;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 3 HVSIVEAWSNDNTPYLHD 20
Db 97 NADLVVLWTDGDTAYPAD 114

RESULT 11
US-10-528-031-1
; Sequence 1, Application US/10528031
; Publication No. US20050262577A1
; GENERAL INFORMATION:
; APPLICANT: ORIDIS BIOMED Forschungs- und Entwicklungs GmbH
; APPLICANT: Guelly, Christian
; APPLICANT: Buck, Charles R.
; APPLICANT: Zatioukal, Kurt
; TITLE OF INVENTION: Polypeptides and nucleic acids encoding these and their use for
; TITLE OF INVENTION: prevention, diagnosis or treatment of liver disorders and epithe
; FILE REFERENCE: Oridis Biomed
; CURRENT APPLICATION NUMBER: US/10/528,031
; CURRENT FILING DATE: 2005-03-16
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-528-031-1

Query Match          35.1%; Score 42.5; DB 6; Length 654;
Best Local Similarity 64.3%; Pred. No. 49;
Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 5 SIV-EAWSNDNTPY 17
Db 251 SIVPOSWSNPPTY 264

RESULT 12
US-11-098-686-10593
; Sequence 10593, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10593
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10593

Query Match          34.7%; Score 42; DB 7; Length 502;
Best Local Similarity 37.5%; Pred. No. 44;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 NNHVSIVEAWSNDNTP 16
Db 186 NPYFALLESWKGNGTP 201

RESULT 13
US-11-052-554A-124
; Sequence 124, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 124
; LENGTH: 1627
; TYPE: PRT
; ORGANISM: Mycoplasma pneumoniae
US-11-052-554A-124

Query Match          34.7%; Score 42; DB 7; Length 1627;
Best Local Similarity 39.3%; Pred. No. 1.5e+02;
Matches 11; Conservative 2; Mismatches 3; Indels 12; Gaps 1;

QY 4 VSVIVEA-----WSNDNTPYLH 19
Db 1214 VSVVEASAYKPTSSGQTQSTNSPYLH 1241

RESULT 14
US-11-052-554A-57
; Sequence 57, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 57
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-57

Query Match          33.9%; Score 41; DB 7; Length 746;
Best Local Similarity 43.8%; Pred. No. 94;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 6 IVEAWSNDNTPYLHDD 21
Db 19 VAQAQEPDTPVSHDD 34

RESULT 15
US-10-510-386-218
; Sequence 218, Application US/10510386
```

```

; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510.386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 218
; LENGTH: 1432
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-218

Query Match      33.9%; Score 41; DB 6; Length 1432;
Best Local Similarity 42.1%; Pred. No. 1.8e+02;
Matches      8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy      2 NHVSIVEANSNDTPYLHD 20
Db      533 NHQETIETVSGSETPLKAD 551

```

Search completed: February 11, 2006, 20:59:02
Job time : 0.119502 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 0.433449 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-42
Perfect score: 121
Sequence: 1 NNHVSIVEAWSNDTPYLHDD 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	100.0	1592	2 A38175	glucosyltransferas
2	110	90.9	1475	2 B33135	gtfB protein precu
3	101	83.5	1375	2 J70345	dextranucrase (EC
4	73	60.3	1431	2 A45866	dextranucrase (EC
5	66	54.5	1365	2 A41483	glucosyltransferas
6	64	52.9	1508	2 T31098	probable dextranu
7	60	49.6	1449	2 T30857	glucosyltransferas
8	60	49.6	1449	2 T30552	glucosyltransferas
9	57	47.1	1577	2 T30858	glucosyltransferas
10	56	46.3	1518	2 A44811	glucosyltransferas
11	55	45.5	1599	2 S22737	glucosyltransferas
12	53	43.8	175	2 C86205	glucosyltransferas
13	49	40.5	396	2 T35024	hypothetical prote
14	48	39.7	331	2 B48445	probable glutathio
15	47	38.8	107	2 S40290	glyceraldhyde-3-p
16	47	38.8	336	1 Q3B840	protein-tyrosine-p
17	47	38.8	1090	2 AG1749	Bglf2 protein - hu
18	47	38.8	2294	2 I67630	glucosidase homolo
19	47	38.8	2450	2 S71625	protein-tyrosine-p
20	47	38.8	2466	2 I67629	protein-tyrosine-p
21	47	38.8	2490	1 A54971	protein-tyrosine-p
22	46	38.0	313	2 S59448	hypothetical prote
23	46	38.0	490	2 H70538	probable ppdK prot
24	46	38.0	601	2 E87028	pyruvate, phospat
25	46	38.0	947	2 E86362	hypothetical prote
26	46	38.0	1100	2 AF1460	alpha-xylosidase a
27	46	38.0	1100	2 AG1097	alpha-xylosidase a
28	45.5	37.6	194	2 H72037	conserved hypotet
29	45.5	37.6	194	2 C86586	CT647 hypothetical

30	45.5	37.6	418	2 D90506	4-aminobutyrate am
31	45	37.2	337	2 C64233	glyceraldhyde-3-p
32	45	37.2	347	2 T48610	hypothetical prote
33	45	37.2	524	2 D82220	conserved hypotet
34	45	37.2	525	2 T40088	RhoGEF domain cont
35	45	37.2	740	2 T15197	hypothetical prote
36	45	37.2	895	2 I54343	dystroglycan - hum
37	45	37.2	1131	2 T16217	hypothetical prote
38	45	37.2	1384	2 T02748	hypothetical prote
39	44.5	36.8	835	2 B64689	site-specific DNA-
40	44	36.4	316	1 A42935	N-acetylmuramoyl-L
41	44	36.4	735	2 T19065	probable protein k
42	44	36.4	1313	2 T29027	hypothetical prote
43	43.5	36.0	279	1 SNPF5K	proteasome endopep
44	43.5	36.0	491	1 D64947	glucose-6-phosphat
45	43.5	36.0	491	2 AB0742	glucose-6-phosphat

ALIGNMENTS

RESULT 1

A38175
glucosyltransferase precursor - Streptococcus sobrinus
C;Species: Streptococcus sobrinus
C;Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004
C;Accession: A38175
R;Abo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A;Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A;Reference number: A38175; MUID:91123227; PMID:1704006
A;Accession: A38175
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1592 <ABO>
A;Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:G217032; PIDN:BAA14241.1; PID
F;1093-1112/Domain: cpl repeat homology <CP1>
F;1222-1241/Domain: cpl repeat homology <CP2>
F;1287-1306/Domain: cpl repeat homology <CP3>
F;1330-1351/Domain: cpl repeat homology <CP4>
F;1352-1371/Domain: cpl repeat homology <CP5>
F;1402-1420/Domain: cpl repeat homology <CP6>
F;1465-1484/Domain: cpl repeat homology <CP7>
F;1513-1532/Domain: cpl repeat homology <CP8>

Query Match 100.0%; Score 121; DB 2; Length 1592;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
|||||
DB 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 2

B33135
gtfB protein precursor - Streptococcus mutans
C;Species: Streptococcus mutans
C;Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C;Accession: B33135; A33128
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A;Reference number: A33135; MUID:87308013; PMID:3040685
A;Accession: B33135
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1475 <SHI>
A;Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:G153639; PIDN
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
submitted to the Protein Sequence Database, September 1990
A;Reference number: A33128
A;Accession: A33128

F:1456-1475/Domain: cpl repeat homology <CPR>

Query Match 45.5%; Score 55; DB 2; Length 1599;
Best Local Similarity 50.0%; Pred. No. 6.3;
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 HVSIIVEAWSNDPTPYLHD 20
DB 494 NISILEAWSNDPPYYNE 511
:::||||| |||:::

RESULT 12
C86205
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: C86205
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
ansen, N.P.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Nature 408, 816-820, 2000
C:A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86205
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-175 <STO>
A:Cross-references: UNIPROT:Q9LMJ8; UNIPARC:UPI00000A10C0; GB:AE005172; NID:g8954041; PI
A:Genetics:
A:Map position: 1

Query Match 43.8%; Score 53; DB 2; Length 175;
Best Local Similarity 42.1%; Pred. No. 0.92;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 NHVSIVEAWSNDPTPYLHD 20
DB 109 NHQEVIDANSRQKPLWTD 127
|||:::|||||

RESULT 13
T35024
probable glutathione-dependent aldehyde dehydrogenase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: T35024
R:Seeger, S.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21565
A:Accession: T35024
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-396 <SEE>
A:Cross-references: UNIPROT:Q9XAM4; UNIPARC:UPI00000DB17B; EMBL:AL079355; PIDN:CAB45580.
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SC0EB:SC4C6.16
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 40.5%; Score 49; DB 2; Length 396;
Best Local Similarity 60.0%; Pred. No. 10;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 VEAWSNDPTPYLHD 21
DB 343 VRRWSDEIVPYLTDD 357
| ||| ||| |||
| ||| ||| |||

RESULT 14
B48445
glyceralddehyde-3-phosphate dehydrogenase (phosphorylating) (EC 1.2.1.12) - Leishmania
C:Species: Leishmania mexicana
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: B48445; S25142
R:Hannaert, V.; Blaauw, M.; Kohl, L.; Allert, S.; Oppendoes, F.R.; Michels, P.A.M.
Mol. Biochem. Parasitol. 55, 115-126, 1992
A:Title: Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-phosphate
A:Reference number: A48445; MUID:93063042; PMID:1435864
A:Accession: B48445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <HAN>
A:Cross-references: UNIPROT:Q01558; UNIPARC:UPI000016BF19; EMBL:X65220; NID:g9552; PIDN
C:Superfamily: glyceraldehyde-3-phosphate dehydrogenase
C:Keywords: oxidoreductase

Query Match 39.7%; Score 48; DB 2; Length 331;
Best Local Similarity 42.1%; Pred. No. 12;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDPTPYLH 19
DB 301 NDHFVKLVSWYDNETGYSH 319
|:| : : |||:
|:| : : |||:

RESULT 15
S40290
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 13 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 06-Mar-1994 #sequence_revision 15-Mar-1996 #text_change 05-Oct-2004
C:Accession: S40290
R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
submitted to the EMBL Data Library, June 1993
A:Description: Assessment of the expression levels of murine protein-tyrosine phosphatase
A:Reference number: S40280
A:Accession: S40290
A:Molecule type: mRNA
A:Residues: 1-107 <HEN>
A:Cross-references: UNIPROT:Q64512; UNIPARC:UPI000016CFDA; EMBL:Z23059; NID:g438155; PI
C:Superfamily: protein-tyrosine phosphatase, non-receptor type 13; GLGF domain homology
C:Keywords: phosphoric monoester hydrolase; tyrosine-specific phosphatase
F:1-107/Domain: protein-tyrosine-phosphatase homology (fragment) <FTP>

Query Match 38.8%; Score 47; DB 2; Length 107;
Best Local Similarity 47.6%; Pred. No. 4.3;
Matches 10; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

QY 3 HVSIIVEAWSNDPTPYLHD 21
DB 68 HSHLNFTAWPDHDPSPQDD 88
|:| : |||:
|:| : |||:

Search completed: February 11, 2006, 19:42:59
Job time : 1.43345 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:57:35 ; Search time 2.83362 Seconds
(without alignments)
5228.676 Million cell updates/sec

Title: US-10-797-821-42
Perfect score: 121
Sequence: 1 NNHVSIVEAWSNDTPYLHDD 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	100.0	1590	2	Q55263_9STRE
2	121	100.0	1590	2	Q59983_9STRE
3	121	100.0	1592	1	GTF2_STRDO
4	121	100.0	1597	1	GTF1_STRDO
5	110	90.9	1476	1	GTF2_STRMU
6	101	83.5	1455	1	GTF3_STRMU
7	76	62.8	1016	2	Q9LCJ7_LEUME
8	76	62.8	1454	2	Q69A94_LEUME
9	73	60.3	1462	1	GTFD_STRMU
10	68	56.2	2835	2	Q8G9Q2_LEUME
11	67	55.4	1575	2	Q9LCH3_STROR
12	67	55.4	1577	2	Q54178_STRGN
13	66	54.5	1338	2	Q9WXJ4_9STRE
14	66	54.5	1365	1	GTF5_STRDO
15	64	52.9	1330	2	Q84CN4_LEUME
16	64	52.9	1477	2	Q9L466_LEUME
17	64	52.9	1508	2	Q9EZHS_LEUME
18	64	52.9	1508	2	Q52224_LEUME
19	61	50.4	1506	2	Q56CX8_9STRE
20	61	50.4	1522	2	Q6TXV4_LEUME
21	61	50.4	1527	2	Q8KRE1_LEUME
22	61	50.4	1527	2	Q9ZAR4_LEUME
23	61	50.4	1554	2	Q8KZL5_9STRE
24	60	49.6	374	2	Q6ZXI9_9LACO
25	60	49.6	1449	2	Q68542_STRSL
26	60	49.6	1449	2	Q55264_STRSL
27	60	49.6	1463	2	Q5SBM6_LACFE
28	60	49.6	1512	2	Q9WXJ5_9STRE
29	58	47.9	1561	2	Q5SBM8_9LACO
30	58	47.9	1595	2	Q5SBM3_LACSK
31	57	47.1	1577	2	Q55265_STRSL

32	56	46.3	1518	2	Q00600_STRSL
33	55	45.5	1442	2	Q7R2S1_GIALA
34	55	45.5	1599	2	Q00599_STRSL
35	55	45.5	2493	2	Q6CMV4_KLJLA
36	53	43.8	175	2	Q9LMJ8_ARATH
37	53	43.8	195	2	Q8LDM8_ARATH
38	52.5	43.4	509	2	Q4WBE0_ASPFU
39	51.5	42.6	767	2	Q4IEN6_GIBZE
40	51	42.1	523	2	Q6CI50_YARLI
41	51	42.1	892	2	Q55VC2_CRYNE
42	51	42.1	892	2	Q5KKW3_CRYNE
43	51	42.1	2057	2	Q9RE05_LEUME
44	50	41.3	395	2	Q4FUX5_9GAMM
45	50	41.3	560	2	Q73WJ6_MYCPA

ALIGNMENTS

RESULT 1
Q55263_9STRE
ID Q55263_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE GTF-I.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus;
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase produced from Streptococcus sobrinus ATCC 33478.";
RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29 (1996).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=9122498; PubMed=1827439;
RA Mooser G., Hefta S.A., Faxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases";
RL J. Biol. Chem. 266:8916-8922 (1991).
DR EMBL; D63570; BAA09792.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; 1GVM.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 100.0%; Score 121; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21

Db 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 2

Q59983_9STRE
ID Q59983_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q59983;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)

```

DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfi;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]_TaxID=1310;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OMZ176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; 1HCX.
DR GO; GO:0047849; F:dextranase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_I; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Glycosyltransferase; Signal; Transferase.
FT SIGNAL 1 38
FT CHAIN 39 1590 Glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 121; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
Db 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 3
GTF2_STRDO
ID_GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]_TaxID=1317;
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=6715 / Serotype G;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-

```

```

CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes both
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 16 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; D90213; BAA14241.1; -; Genomic_DNA.
DR HSSP; P06653; 1GVM.
DR InterPro; IPR002479; Cell_wall_bd_put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_I; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38
FT CHAIN 39 1592
FT REPEAT 1093 1142 1.
FT REPEAT 1158 1207 2.
FT REPEAT 1222 1272 3.
FT REPEAT 1287 1337 4.
FT REPEAT 1402 1451 5.
FT REPEAT 1514 1563 6.
FT REPEAT 1577 1592 7 (incomplete).
FT REGION 39 1044 Catalytic (approximate).
FT REGION 1093 1592 7 X tandem repeats.
FT REGION 1093 1592 Glucan-binding (approximate).
SQ SEQUENCE 1592 AA; 176168 MW; BC0A66D079351ECF CRC64;

Query Match 100.0%; Score 121; DB 1; Length 1592;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDTPYLHDD 21
Db 478 NNHVSIVEAWSNDTPYLHDD 498

RESULT 4
GTF1_STRDO
ID_GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfi;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]_TaxID=1317;
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RA "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
RT sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-

```


[1] NUCLEOTIDE SEQUENCE.
 RN STRAIN=GS-5; PubMed=2148600;
 RC MEDLINE=91100958; PubMed=2148600;
 RX Honda O., Kato C., Kuramitsu H.K.;
 RA "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
 RT the glucosyltransferase-S enzyme."; J. Gen. Microbiol. 136:2099-2105 (1990).
 RL [2]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
 RC MT4467 / Serotype e, and MT8148 / Serotype c;
 RX MEDLINE=98311643; PubMed=9570124;
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
 RA Kimura S., Hamada S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RL Streptococcus mutans."; J. Gen. Microbiol. Lett. 161:331-336 (1998).
 RN [3]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
 RX MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
 RL pathogen."; Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to
 CC play a key role in the development of the dental plaque because of
 CC their ability to adhere to smooth surfaces and mediate the
 CC aggregation of bacterial cells and food debris.
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
 CC fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
 CC forms of glucans.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -!- SIMILARITY: Contains 6 cell wall binding repeats.
 CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC EMBL; M29296; AAA26895.1; -; Genomic DNA.
 CC EMBL; D88653; BAA26103.1; -; Genomic DNA.
 CC EMBL; D88656; BAA26107.1; -; Genomic DNA.
 CC EMBL; D88659; BAA26111.1; -; Genomic DNA.
 CC EMBL; D88662; BAA26115.1; -; Genomic DNA.
 CC EMBL; D89979; BAA26121.1; -; Genomic DNA.
 CC EMBL; A014932; AAN58619.1; -; Genomic DNA.
 CC HSSP; P06653; 1GWM.
 DR InterPro; IPR002479; Cell wall bd put.
 DR Pfam; PF01473; Glyco_hydro_70.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal;
 KW Transferase.
 FT SIGNAL 1 ? Potential.
 FT CHAIN ? 1462 Glucosyltransferase-S.
 FT REPEAT 1232 1295 1.
 FT REPEAT 1296 1359 2.
 FT REPEAT 1360 1423 3.
 FT REGION 1232 1423 3 X 63 AA approximate tandem repeats.
 FT VARIANT 10 10 Y -> H (in strain GS-5, strain MT4239,
 FT strain MT4245, strain MT4251, strain
 FT MT4467 and strain MT8148).
 FT VARIANT 19 19 I -> V (in strain GS-5, strain MT4239,

FT strain MT4245, strain MT4251, strain
 FT MT4467 and strain MT8148).
 FT K -> E (in strain MT4467).
 FT A -> S (in strain MT4239 and strain
 FT MT4245).
 FT A -> T (in strain MT4251 and strain
 FT MT8148).
 FT T -> I (in strain MT4239 and strain
 FT MT4245).
 FT A -> V (in strain MT4239, strain MT4245
 FT and strain MT8148).
 FT A -> S (in strain GS-5 and strain
 FT MT4467).
 FT A -> V (in strain MT4245).
 FT A -> T (in strain GS-5, strain MT4239,
 FT strain MT4245, strain MT4251, strain
 FT MT4467 and strain MT8148).
 FT V -> L (in strain MT4239).
 FT D -> N (in strain MT8148).
 FT E -> D (in strain MT4239, strain MT4245
 FT and strain MT4251).
 FT D -> N (in strain MT4239, strain MT4245
 FT and strain MT4251).
 FT Q -> H (in strain MT4245).
 FT D -> N (in strain MT4239 and strain
 FT MT4251).
 FT E -> K (in strain MT4239).
 FT V -> F (in strain MT4239).
 FT F -> L (in strain MT4239, strain MT4251
 FT and strain MT4467).
 FT KKKYQ -> EKEYTL (in strain MT4251).
 FT A -> S (in strain MT4239).
 FT TDQGSSEA -> ADKGNDS (in strain MT4251).
 FT TDQGS -> ADKGN (in strain MT4239 and
 FT strain MT4245).
 FT T -> A (in strain GS-5, strain MT4239,
 FT strain MT4245, strain MT4251, strain
 FT MT4467 and strain MT8148).
 FT D -> Y (in strain MT4251).
 FT E -> K (in strain MT4245 and strain
 FT MT4251).
 FT LG -> IR (in strain MT4251).
 FT G -> R (in strain MT4245).
 FT G -> R (in strain MT4239).
 FT H -> Q (in strain GS-5).
 FT S -> N (in strain MT4239).
 FT Y -> C (in strain MT4251 and strain
 FT MT4467).
 FT F -> L (in strain MT4467).
 FT Q -> P (in strain MT4245).
 FT K -> T (in strain MT4245).
 FT N -> D (in strain MT4245).
 FT G -> D (in strain GS-5 and strain
 FT MT4467).
 FT G -> R (in strain GS-5).
 FT R -> K (in strain MT4467).
 FT RYDNGSMVNVKVVTLANGRRIGIDRWGIARYY -> VY
 FT R (in Ref. 1).
 SQ SEQUENCE 1462 AA; 163388 MW; CE4A279C4D708645 CRC64;
 Query Match 60.3%; Score 73; DB 1; Length 1462;
 Best Local Similarity 68.4%; Pred. No. 0.057;
 Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 2 NHVSIVEAWSNDTPYLHD 20
 Db 497 NHLSLLEAWSNDPQYNKD 515
 ||:|||||
 ||:|||||
 RESULT 10
 Q8G9Q2 LEUME
 ID Q8G9Q2 LEUME PRELIMINARY; PRT; 2835 AA.
 AC Q8G9Q2;

RESULT	11
ID	Q9LCH3 STOR PRELIMINARY; PRT; 1575 AA.
AC	Q9LCH3;
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE	01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE	Glucoyltransferase.
GN	Name=gtfr;
OS	Streptococcus oralis.
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=1303;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RC	STRAIN=ATCC10557;
RX	MEDLINE=202311779; PubMed=10768934;
RX	DOI=10.1128/JAI.68.5.2475-2483.2000;
RA	Fujikawa T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RA	"Purification, characterization, and molecular analysis of the gene
RT	encoding glucoyltransferase from Streptococcus oralis.";
RL	Infect. Immun. 68:2475-2483(2000).
RL	EMBL; AB025228; BAA95201.1; -; Genomic_DNA.
DR	HSP; P06653; 1HCX.
DR	GO; GO:0016740; F:transferase activity; IEA.
DR	GO; GO:0009250; P:glucan biosynthesis; IEA.
DR	InterPro; IPR002479; CW_binding_70.
DR	InterPro; IPR003318; Glyco_hydro_70.
DR	Pfam; PF01473; CW_binding_1; 4.
DR	Pfam; PF02324; Glyco_hydro_70; 1.
KW	Transferase.
SQ	SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2B543 CRC64;
Query Match	55.4%; Score 67; DB 2; Length 1575;
Best Local Similarity	66.7%; Pred. NO. 0.54;

```

RESULT 13
ID Q9WXJ4_9STRE PRELIMINARY; PRT; 1338 AA.
AC Q9WXJ4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GTF-S.
GN Name=gtfS;
OS Streptococcus criceti.
OC Plasmid pAM1.
OC OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxId=1333;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RL Inoue M., Fukui K., Miyagi A.;
RR Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

```

```

FT REPEAT 1150 1169 Cell wall binding 5.
FT REPEAT 1170 1190 Cell wall binding 6.
FT REPEAT 1225 1243 Cell wall binding 7.
FT REPEAT 1289 1308 Cell wall binding 8.
FT REPEAT 1309 1328 Cell wall binding 9.
FT REPEAT 1331 1352 Cell wall binding 10.
FT REGION 198 1061 Catalytic (approximate).
SQ SEQUENCE 1365 AA; 151591 MW; 167296BSA2E8C476 CRC64;

Query Match 54.5%; Score 66; DB 1; Length 1365;
Best Local Similarity 57.9%; Pred. No. 0.65;
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 NHVSIVEAWSNDNTPYLHD 20
    :|||:|||||
Db 469 DHLSILEANGSNDNDYVKD 487

RESULT 15
Q84CN4 LEUME
ID Q84CN4 LEUME PRELIMINARY; PRT; 1330 AA.
AC Q84CN4;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Dextranucrase DsrR (BC 2.4.1.5).
GN Name=dsrR;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
NCBI_TaxID=1245;
RX [1]_
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=NRRL B-1501;
RC Kim C.H., Moon J.O., Jang E.K.;
RA Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY142210; AA38835.1; -; Genomic DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 1.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1330 AA; 148863 MW; D945CBB36CF75797 CRC64;

Query Match 52.9%; Score 64; DB 2; Length 1330;
Best Local Similarity 55.0%; Pred. No. 1.3;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 NNHVSIVEAWSNDNTPYLHD 20
    |||:|||||
Db 386 NQHLSILEDSHNDPYVKD 405

Search completed: February 11, 2006, 19:39:14
Job time : 3.83362 secs

```

Search completed: February 11, 2006, 19:39:14
Job time : 3.83362 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 18:56:56 ; Search time 1.99074 Seconds
(without alignments)
4414.224 Million cell updates/sec

Title: US-10-797-821-43

Perfect score: 105
Sequence: 1 VVIANNVDFVSWGITDFEM 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003s:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	20	ADX37286	Adx37286 Streptoco
2	105	100.0	1590	ADD93657	Add93657 Streptoco
3	105	100.0	1590	ADX37280	Adx37280 Streptoco
4	105	100.0	1592	AAR32925	Aar32925 Glucosylt
5	90	85.7	1017	AAU79285	Aau79285 Streptoco
6	90	85.7	1375	AAU98028	Aau98028 S. mutans
7	90	85.7	1375	AAU79288	Aau79288 Streptoco
8	90	85.7	1375	ADD93655	Add93655 Streptoco
9	90	85.7	1375	ADX37278	Adx37278 Streptoco
10	90	85.7	1475	AAU98031	Aau98031 S. mutans
11	90	85.7	1475	AAU98036	Aau98036 S. mutans
12	90	85.7	1475	AAU98037	Aau98037 S. mutans
13	90	85.7	1475	AAU98040	Aau98040 S. mutans
14	90	85.7	1475	AAU98035	Aau98035 S. mutans
15	90	85.7	1475	AAU98033	Aau98033 S. mutans
16	90	85.7	1475	AAU98034	Aau98034 S. mutans
17	90	85.7	1475	AAU98030	Aau98030 S. mutans
18	90	85.7	1475	AAU98039	Aau98039 S. mutans
19	90	85.7	1475	AAU98027	Aau98027 S. mutans
20	90	85.7	1475	AAU98032	Aau98032 S. mutans
21	90	85.7	1475	AAU98038	Aau98038 S. mutans
22	90	85.7	1475	ADD93654	Add93654 Streptoco
23	90	85.7	1475	ADX37277	Adx37277 Streptoco
24	90	85.7	1476	AAU79284	Aau79284 Streptoco

25	72	68.6	855	5	ABB98573	Abb98573 Dextran s
26	72	68.6	2835	5	ABB98574	Abb98574 Dextran s
27	72	68.6	2835	6	ABR55594	Abrr55594 Amino aci
28	71	67.6	1518	7	ADD93660	Add93660 Streptoco
29	71	67.6	1518	9	ADX37283	Adx37283 Streptoco
30	69	65.7	1365	7	ADD93659	Add93659 Streptoco
31	69	65.7	1365	9	ADX37282	Adx37282 Streptoco
32	69	65.7	1499	7	ADC54806	Adc54806 Protein S
33	68	64.8	1527	5	AAU80055	Aau80055 Leuconost
34	68	64.8	1527	7	ADC54807	Adc54807 Leuconost
35	65	61.9	1430	5	AAU98043	Aau98043 S. mutans
36	65	61.9	1430	5	AAU98044	Aau98044 S. mutans
37	65	61.9	1430	5	AAU98045	Aau98045 S. mutans
38	65	61.9	1430	5	AAU98042	Aau98042 S. mutans
39	65	61.9	1430	5	AAU98029	Aau98029 S. mutans
40	65	61.9	1430	5	AAU98041	Aau98041 S. mutans
41	65	61.9	1430	7	ADD93656	Add93656 Streptoco
42	65	61.9	1430	9	ADX37279	Adx37279 Streptoco
43	65	61.9	2147	6	ABR63231	Abrr63231 Glucansuc
44	63	60.0	2055	6	ABR63235	Abrr63235 Glucansuc
45	56	53.3	1577	2	AAR91047	Aarr91047 Alpha-D-g

ALIGNMENTS

RESULT 1
ADX37286
ID ADX37286 standard; peptide; 20 AA.
XX
AC ADX37286;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus sobrinus glukan binding protein B peptide #2.
XX
KW immunogenicity; immune stimulation; glukan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus sobrinus.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PE 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2003US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
XX (SMIT/) SMITH D J.
XX (TAUB/) TAUBMAN M A.
XX Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glukan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.
XX
XX Example 5; SEQ ID NO 43; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glukan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a

CC Streptococcus sobrinus GbpB-derived peptide of the invention.

XX Sequence 20 AA;

Query Match 100.0%; Score 105; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 VVIANNVDFVSWGITDFEM 20

RESULT 2

ADD93657
ID ADD93657 standard; protein; 1590 AA.

XX

AC ADD93657;

XX 29-JAN-2004 (first entry)

DE Streptococcus sobrinus glucosyltransferase-I.

KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

XX Streptococcus sobrinus.

XX WO2003075845-A2.

XX 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US006962.

XX 07-MAR-2002; 2002US-0363209P.

XX 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

XX Claim 16; Page 14; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus sobrinus glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutants glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

XX Sequence 1590 AA;

Query Match 100.0%; Score 105; DB 7; Length 1590;
Best Local Similarity 100.0%; Pred. No. 4.1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
| | | | | | | | | | | | | | | | | | | | | |
Db 847 VVIANNVDFVSWGITDFEM 866

RESULT 3

ADX37280

ID ADX37280 standard; protein; 1590 AA.

XX

AC ADX37280;

XX 21-APR-2005 (first entry)

XX Streptococcus sobrinus glucan binding protein B.

KW immunogenicity; immune stimulation; glucan binding protein-B; microparticle; major histocompatibility complex; tooth disease.

XX Streptococcus sobrinus.

XX US2005031633-A1.

XX 10-FEB-2005.

XX 09-MAR-2004; 2004US-00797821.

XX 13-APR-1998; 98US-0081550P.

XX 08-JAN-1999; 99US-0115142P.

XX 12-APR-1999; 99US-00290049.

XX 07-MAR-2002; 2002US-0363209P.

XX 08-AUG-2002; 2002US-0402483P.

XX 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.

XX (TAUB/) TAUBMAN M A.

XX Smith DJ, Taubman MA;

XX WPI; 2005-151644/16.

XX New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.

XX Claim 7; SEQ ID NO 37; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a Streptococcus sobrinus GbpB protein of the invention.

XX Sequence 1590 AA;

Query Match 100.0%; Score 105; DB 9; Length 1590;
Best Local Similarity 100.0%; Pred. No. 4.1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
| | | | | | | | | | | | | | | | | | | | | |
Db 847 VVIANNVDFVSWGITDFEM 866

RESULT 4

AAR32925
ID AAR32925 standard; protein; 1592 AA.

XX

AC AAR32925;

XX 28-JUN-1993 (first entry)

XX Glucosyltransferase I.

XX GT-1; Streptococcus; dental; caries.

XX Streptococcus sobrinus.

PN JP05023188-A.
 XX 02-FEB-1993.
 XX 25-JUL-1991; 91JP-00186592.
 XX 25-JUL-1991; 91JP-00186592.
 XX (KATO/) KATO K.
 XX (FUKU/) FUKUI I.
 XX WPI; 1993-079449/10.
 XX N-PSDB; AAQ37760.
 XX DNA sequence Glucosyltransferase-I - comprises Streptococcus sobrinus
 PT DNA sequence with at least one nucleotide added or deleted.
 XX Claim 13; Page 15; 29pp; Japanese.
 XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating S.
 CC sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially
 CC digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp
 CC fragment was ligated into pUC18 and E. coli JM109 transformed with it. A
 CC GT-1 expressing clone was isolated and sequenced. The clone may be used
 CC in the development of a drug for dental caries
 XX Sequence 1592 AA;
 SQ

Query Match 100.0%; Score 105; DB 2; Length 1592;
 Best Local Similarity 100.0%; Pred. No. 4.1e-07;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VVIANNVDFVSMGTTDFEM 20
 |||||
 DB 847 VVIANNVDFVSMGTTDFEM 866
 |||||

RESULT 5
 AAU79285
 ID AAU79285 standard; protein; 1017 AA.
 XX AAU79285;
 AC
 XX 13-AUG-2002 (first entry)
 DT
 XX Streptococcus mutans monoclonal antibody-related protein #2.
 DE
 XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;
 KW glucosyl transferase-B; immunotherapy.
 XX Streptococcus mutans.
 OS
 XX JP2002114709-A.
 PN
 XX 16-APR-2002.
 PD
 XX 04-OCT-2000; 2000JP-00304889.
 PF
 XX 04-OCT-2000; 2000JP-00304889.
 PR
 XX (UYNI-) UNIV NIPPON.
 PA
 XX WPI; 2002-448885/48.
 DR
 XX Anti-carries agent composed of a monoclonal antibody against an inhibitory
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-
 PT B (GTF-B) of Streptococcus mutans.
 XX
 XX Claim 4; Page 17-19; 28pp; Japanese.
 PS
 XX The invention relates to a monoclonal antibody against dental caries and
 CC an anti-carries agent composed of a monoclonal antibody produced by

CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl
 CC transferase-B. The monoclonal antibody specifically inhibits water
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
 CC transferase-B and is used in the immunotherapy of dental caries. This
 CC sequence represents a Streptococcus mutans monoclonal antibody-related
 CC protein
 XX
 XX Sequence 1017 AA;
 SQ

Query Match 85.7%; Score 90; DB 5; Length 1017;
 Best Local Similarity 80.0%; Pred. No. 5.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VVIANNVDFVSMGTTDFEM 20
 |||||
 DB 816 VVIANNVDFVSMGTTDFEM 835
 |||||

RESULT 6
 AAU98028
 ID AAU98028 standard; protein; 1375 AA.
 XX AAU98028;
 AC
 XX 27-AUG-2002 (first entry)
 DT
 XX S. mutans glucosyltransferase GTFC.
 DE
 XX Glucosyltransferase; GTFC; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture.
 XX Streptococcus mutans.
 OS
 XX US2002031826-A1.
 PN
 XX 14-MAR-2002.
 PD
 XX 19-DEC-2000; 2000US-00740274.
 PF
 XX 07-JUN-1995; 95US-00478704.
 PR
 XX 07-JUN-1995; 95US-00482711.
 PR
 XX 07-JUN-1995; 95US-00485243.
 PR
 XX 16-JAN-1998; 98US-00007999.
 PR
 XX 16-JAN-1998; 98US-00088172.
 PR
 XX 20-JAN-1998; 98US-00009620.
 PR
 XX 11-DEC-1998; 98US-00210361.
 XX (NICH/) NICHOLS S E.
 XX Nichols SE;
 XX WPI; 2002-414332/44.
 DR
 XX N-PSDB; ABK52939.
 DR
 XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 XX
 XX Disclosure; Page 30-33; 44pp; English.
 PS
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 1448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes PI or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host

CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GFP, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GFP, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents GFPc
 XX
 SQ Sequence 1375 AA;

Query Match 85.7%; Score 90; DB 5; Length 1375;

Best Local Similarity 80.0%; Pred. No. 7.8e-05;

Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20
 |||| ||||| ||:|||||
 Db 876 VVIANKVDKFAEWGVTDFEM 895

RESULT 7

AAU79288

ID AAU79288 standard; protein; 1375 AA.

AC AAU79288;

DT 13-AUG-2002 (first entry)

DE Streptococcus mutans monoclonal antibody-related protein #5.

XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;

KW glucosyl transferase-B; immunotherapy.

XX Streptococcus mutans.

OS JP2002114709-A.

PN 16-APR-2002.

PD 04-OCT-2000; 2000JP-00304889.

PF 04-OCT-2000; 2000JP-00304889.

PR (UYNI-) UNIV NIPPON.

PA WPI; 2002-448895/48.

XX Anti-carries agent composed of a monoclonal antibody against an inhibitory

PT enzyme against water insoluble glucan synthetase of glucosyl transferase-

PT B (GTF-B) of Streptococcus mutans.

XX Disclosure; Page 22-25; 28pp; Japanese.

XX The invention relates to a monoclonal antibody against dental caries and

CC an anti-carries agent composed of a monoclonal antibody produced by

CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)

CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl
 CC transferase-B. The monoclonal antibody specifically inhibits water
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
 CC transferase-B and is used in the immunotherapy of dental caries. This
 CC sequence represents a Streptococcus mutans monoclonal antibody-related
 CC protein
 XX
 SQ Sequence 1375 AA;

Query Match 85.7%; Score 90; DB 5; Length 1375;

Best Local Similarity 80.0%; Pred. No. 7.8e-05;

Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20
 |||| ||||| ||:|||||
 Db 876 VVIANKVDKFAEWGVTDFEM 895

RESULT 8

ADD93655

ID ADD93655 standard; protein; 1375 AA.

AC ADD93655;

DT 29-JAN-2004 (first entry)

DE Streptococcus mutans glucosyltransferase-C.

XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

PN WO2003075845-A2.

PD 18-SEP-2003.

PF 07-MAR-2003; 2003WO-US006962.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX Composition useful as vaccines for dental caries comprises a fragment of
 PT a glucan binding protein-B binding to a major histocompatibility complex
 PT class II protein.

XX Claim 16; Page 13; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus mutans
 CC glucosyltransferase-C (GTF-C). Peptide fragments of GTF-C, especially
 CC from the catalytic domain of the polypeptide, can be used in immunogenic
 CC compositions and subunit vaccines for dental caries. These compositions
 CC comprise a major histocompatibility complex (MHC) class II protein-
 CC binding peptide from S. mutans glucan binding protein-B (GbpB)
 CC covalently linked with a peptide fragment of a streptococcal
 CC glucosyltransferase. The compositions are used in a claimed method of
 CC eliciting production of an antibody in a mammal. Dieptopic or
 CC multipitopic polypeptides can be prepared synthetically or by
 CC recombinant DNA technology. Antibodies raised against MHC class II
 CC binding fragments of GbpB can be used in passive immunisation.

XX Sequence 1375 AA;

Query Match 85.7%; Score 90; DB 7; Length 1375;

Best Local Similarity 80.0%; Pred. No. 7.8e-05;

Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVIANNVDFKFSWGVTDFEM 20
 ||||| ||||| ||||| |||||
 Db 876 VVIANNVDFKFAEWGVTDFEM 895

RESULT 9

ADK37278
 ID -ADK37278 standard; protein; 1375 AA.

XX AC ADK37278;
 XX 21-APR-2005 (first entry)
 XX Streptococcus mutant glucan binding protein B variant #7.
 DE immunogenicity; immune stimulation; glucan binding protein-B;
 KW microparticle; major histocompatibility complex; tooth disease.
 XX Streptococcus mutans.
 OS Streptococcus mutans.

XX US2005031633-A1.

XX 10-FEB-2005.

XX 09-MAR-2004; 2004US-00797821.

XX 13-APR-1998; 98US-0081550P.

XX 08-JAN-1999; 99US-0115142P.

XX 12-APR-1999; 99US-00290049.

XX 07-MAR-2002; 2002US-0363209P.

XX 08-AUG-2002; 2002US-0402483P.

XX 07-MAR-2003; 2003US-00383930.

XX (SMITH/) SMITH, J.
 XX (TAUB/) TAUBMAN M. A.

XX Smith DJ, Taubman MA;
 XX WPI; 2005-151644/16.

XX New composition comprising a fragment of a glucan binding protein-B
 PT (GbpB) that binds to MHC class II protein, and a biocompatible
 PT microparticle, useful for producing an antibody (claimed) for immunizing
 PT mammals against dental caries.

XX Claim 7; SEQ ID NO 35; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the
 CC fragment binds to a major histocompatibility complex (MHC) class II
 CC protein. The composition is useful for producing an antibody for
 CC immunizing mammals against dental caries. This sequence corresponds to a
 CC Streptococcus mutans GbpB protein of the invention.

XX Sequence 1375 AA;

Query Match 85.7%; Score 90; DB 9; Length 1375;
 Best Local Similarity 80.0%; Pred. No. 7.8e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVIANNVDFKFSWGVTDFEM 20
 ||||| ||||| ||||| |||||
 Db 876 VVIANNVDFKFAEWGVTDFEM 895

RESULT 10

AAU98031
 ID AAU98031 standard; protein; 1475 AA.

XX AC AAU98031;

XX 27-AUG-2002 (first entry)

XX

DE S. mutans glucosyltransferase GTFB mutant D457N.
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 KW coating composition; glucan; starch; latex; thermoplastic molecule;
 KW amyloplast; vacuole; paper manufacture; mutant; mutin.
 XX Streptococcus mutans.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

XX 07-JUN-1995; 95US-00485243.

XX 16-JAN-1998; 98US-00007999.

XX 16-JAN-1998; 98US-00008172.

XX 20-JAN-1998; 98US-00009620.

XX 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S. E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from 148V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present

CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36

XX SQ Sequence 1475 AA;

Query Match 85.7%; Score 90; DB 5; Length 1475;
 Best Local Similarity 80.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGVTDFEM 20
 ||||| ||||| ||:|||||
 Db 849 VVIANNVDKFAEWGVTDFEM 868

RESULT 11
 AAU98036
 ID AAU98036 standard; protein; 1475 AA.

XX AC AAU98036;

XX DT 27-AUG-2002 (first entry)

XX DE S. mutans glucosyltransferase GTFB mutant D567T/D571K.

XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 XX coating composition; glucan; starch; latex; thermoplastic molecule;
 XX amyloplast; vacuole; paper manufacture; mutant; mutein.

XX OS Streptococcus mutans.
 XX OS Synthetic.

XX FH Key Location/Qualifiers

FT FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"

FT FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"

XX US2002031826-A1.

XX PD 14-MAR-2002.

XX PF 19-DEC-2000; 2000US-00740274.

XX PR 07-JUN-1995; 95US-00478704.

XX PR 07-JUN-1995; 95US-00482711.

XX PR 07-JUN-1995; 95US-00485243.

XX PR 16-JAN-1998; 98US-00007999.

XX PR 16-JAN-1998; 98US-00008172.

XX PR 20-JAN-1998; 98US-00009620.

XX PR 11-DEC-1998; 98US-00210361.

XX PA (NICH/) NICHOLS S E.

XX PI Nichols SE;

XX XX WPI; 2002-414332/44.

XX XX Glucosyltransferase B or D protein useful for producing a glucan useful

XX as substitutes for and additions to modified starch and latexes in paper

XX manufacture, comprises mutations in specific positions.

XX PS Claim 36; Page; 44pp; English.

XX XX The invention an isolated protein comprising a glucosyltransferase (GTF)

XX B polypeptide having changes at position from I48V, D457N, D567T,

XX KI014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/KI014T,

CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.

CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36

XX XX SQ Sequence 1475 AA;

Query Match 85.7%; Score 90; DB 5; Length 1475;
 Best Local Similarity 80.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGVTDFEM 20
 ||||| ||||| ||:|||||
 Db 849 VVIANNVDKFAEWGVTDFEM 868

RESULT 12
 AAU98037
 ID AAU98037 standard; protein; 1475 AA.

XX AC AAU98037;

XX DT 27-AUG-2002 (first entry)

XX DE S. mutans glucosyltransferase GTFB mutant D567T/D571K/KI014T.

XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
 XX coating composition; glucan; starch; latex; thermoplastic molecule;
 XX amyloplast; vacuole; paper manufacture; mutant; mutein.

XX OS Streptococcus mutans.
 XX OS Synthetic.

XX FH Key Location/Qualifiers

FT FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"

FT FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"

FT FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"

XX US2002031826-A1.

XX PD 14-MAR-2002.

XX PF 19-DEC-2000; 2000US-00740274.

CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GFP, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GFPB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GFPB sequence appearing as AAU98027 and the information in
 CC claim 36
 XX
 SQ Sequence 1475 AA;

Query Match 85.7%; Score 90; DB 5; Length 1475;
 Best Local Similarity 80.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGVTDFEM 20
 ||||| ||||| ||:|||||
 Db 849 VVIANNVDFKFAEWGVTDFEM 868

RESULT 14

AAU98035
 ID AAU98035 standard; protein; 1475 AA.

AC AAU98035;

XX 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant D457N/D571K.

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"

FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"

FT US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from I448V, D457N, D567T,

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,

CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GFP mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GFP mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant comprising the
 CC cell, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GFP, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GFP, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 XX
 SQ Sequence 1475 AA;

Query Match 85.7%; Score 90; DB 5; Length 1475;

Best Local Similarity 80.0%; Pred. No. 8.5e-05;

Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGVTDFEM 20

||||| ||||| ||:|||||

Db 849 VVIANNVDFKFAEWGVTDFEM 868

RESULT 15

AAU98033

ID AAU98033 standard; protein; 1475 AA.

AC AAU98033;

DT 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant K1014T.

KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"

FT US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.
 PR 07-JUN-1995; 95US-00482711.
 PR 07-JUN-1995; 95US-00485243.
 PR 16-JAN-1998; 98US-00007999.
 PR 16-JAN-1998; 98US-00008172.
 PR 20-JAN-1998; 98US-00009620.
 PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.
 PA
 XX Nichols SE;
 PI
 XX WPI; 2002-414332/44.
 DR
 XX

XX Glucosyltransferase B or D protein useful for producing a glucan useful
 PT as substitutes for and additions to modified starch and latexes in paper
 PT manufacture, comprises mutations in specific positions.
 PT
 XX
 PS Claim 36; Page; 44pp; English.
 PS
 XX

XX The invention an isolated protein comprising a glucosyltransferase (GTF)
 CC B polypeptide having changes at position from I448V, D457N, D567T,
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the
 CC GTF mutant, an expression cassette comprising the polynucleotide operably
 CC linked to a promoter, a vector comprising the expression cassette, host
 CC cell introduced with the vector, a transgenic plant, a paper sizing and/or
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
 CC coating composition comprising a glucan produced in a plant transformed
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
 CC thermoplastic molecule or their combinations or glucan and starch where
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line
 CC deficient in starch biosynthesis, transformed with a gene encoding a
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
 CC comprising the glucan (paper sizing/coating agent). The vector is useful
 CC for producing a glucan in a plant. The method comprises transforming a
 CC plant cell with the vector, growing the plant cell under plant growing
 CC conditions to produce a regenerated plant and inducing expression of the
 CC polynucleotide, for a time sufficient to produce the glucan in the
 CC regenerated plant, where the vector contains a transit sequence from
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.
 CC Glucans are useful as substitutes for and additions to modified starch
 CC and latexes in paper manufacture. Unlike prior art techniques, which
 CC require input materials that produce chemical effluents, paper
 CC manufacture utilising the glucan produced by GTF, which utilises
 CC biologically produced input materials, is more cost-effective and
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
 CC properties and impart gloss to the paper during coating step. The present
 CC sequence represents a GTFB mutant of the invention. Note: The present
 CC sequence is not shown in the specification but was created by the indexer
 CC using the GTFB sequence appearing as AAU98027 and the information in
 CC claim 36
 XX

SQ Sequence 1475 AA;

Query Match 85.7%; Score 90; DB 5; Length 1475;
 Best Local Similarity 80.0%; Pred. No. 8.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVVKFVSGVITDFEM 20
 ||||| ||||| ||||| |||||
 Db 849 VVIANNVVKFVSGVITDFEM 868

Search completed: February 11, 2006, 19:15:28
 Job time : 1.99074 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:39:30 ; Search time 0.437886 Seconds
(without alignments)
3776.130 Million cell updates/sec

Title: US-10-797-821-43

Perfect score: 105
Sequence: 1 VVIANNVDKFSWGVTDFEM 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/PCUS_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/RS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfileseq.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	85.7	1375	2	US-09-210-361-4
2	90	85.7	1375	2	US-09-740-274-4
3	90	85.7	1475	2	US-09-007-999-2
4	90	85.7	1475	2	US-09-210-361-2
5	90	85.7	1475	2	US-09-740-274-2
6	68	64.8	522	2	US-09-995-749A-11
7	68	64.8	523	2	US-09-604-957-5
8	65	61.9	545	2	US-09-604-957-4
9	65	61.9	545	2	US-09-995-749A-10
10	65	61.9	1430	2	US-09-008-172-2
11	65	61.9	1430	2	US-09-210-361-6
12	65	61.9	1430	2	US-09-740-274-6
13	56	53.3	1577	1	US-08-793-824-2
14	53	50.5	535	2	US-09-604-957-7
15	53	50.5	535	2	US-09-995-749A-13
16	53	50.5	1278	2	US-09-604-957-3
17	53	50.5	1781	2	US-09-995-749A-2
18	49	46.7	659	2	US-09-583-110-4043
19	49	46.7	661	2	US-09-107-433-3700
20	46	43.8	402	2	US-09-134-001C-4138
21	45.5	43.3	399	1	US-08-282-197C-54
22	45.5	43.3	837	2	US-09-390-234-14
23	45.5	43.3	837	2	US-09-603-311-14
24	45	42.9	584	2	US-09-604-957-6
25	45	42.9	584	2	US-09-995-749A-12
26	45	42.9	2057	2	US-09-499-203-2
27	44	41.9	810	1	US-08-785-271-7

28	43	41.0	102	1	US-08-480-473B-47	Sequence 47, Appl
29	43	41.0	102	2	US-08-915-213-47	Sequence 47, Appl
30	43	41.0	102	2	US-09-235-217-47	Sequence 47, Appl
31	43	41.0	202	2	US-08-949-155-2	Sequence 2, Appl
32	43	41.0	202	2	US-09-819-964-2	Sequence 2, Appl
33	43	41.0	212	2	US-09-328-352-6684	Sequence 6684, Ap
34	43	41.0	245	2	US-09-438-833-3	Sequence 3, Appl
35	43	41.0	330	2	US-09-438-833-4	Sequence 4, Appl
36	43	41.0	373	1	US-08-480-473B-3	Sequence 3, Appl
37	43	41.0	373	2	US-08-915-213-3	Sequence 3, Appl
38	43	41.0	373	2	US-09-235-217-3	Sequence 3, Appl
39	43	41.0	373	4	PCT-US96-10251-3	Sequence 3, Appl
40	43	41.0	613	2	US-09-438-833-6	Sequence 6, Appl
41	43	41.0	623	2	US-09-967-388-2	Sequence 2, Appl
42	43	41.0	652	2	US-09-438-833-5	Sequence 5, Appl
43	43	41.0	756	2	US-09-438-833-11	Sequence 11, Appl
44	43	41.0	805	1	US-08-480-473B-4	Sequence 4, Appl
45	43	41.0	805	2	US-08-915-213-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4
Query Match 85.7%; Score 90; DB 2; Length 1375;
Best Local Similarity 80.0%; Pred. No. 2e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 VVIANNVDKFSWGVTDFEM 20
DB 876 VVIANNVDKFAEWGVTDFEM 895
RESULT 2
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19

```
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4
```

```
Query Match 85.7%; Score 90; DB 2; Length 1375;
Best Local Similarity 80.0%; Pred. No. 2e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDFKFSWGVTDFEM 20
||||| ||||| |||||
Db 876 VVIANNVDFKFAEWGVTDFEM 895
```

```
RESULT 3
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2
```

```
Query Match 85.7%; Score 90; DB 2; Length 1475;
Best Local Similarity 80.0%; Pred. No. 2.1e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDFKFSWGVTDFEM 20
||||| ||||| |||||
Db 849 VVIANNVDFKFAEWGVTDFEM 868
```

```
RESULT 4
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
```

```
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
```

```
Query Match 85.7%; Score 90; DB 2; Length 1475;
Best Local Similarity 80.0%; Pred. No. 2.1e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDFKFSWGVTDFEM 20
||||| ||||| |||||
Db 849 VVIANNVDFKFAEWGVTDFEM 868
```

```
RESULT 5
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2
```

```
Query Match 85.7%; Score 90; DB 2; Length 1475;
Best Local Similarity 80.0%; Pred. No. 2.1e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDFKFSWGVTDFEM 20
||||| ||||| |||||
Db 849 VVIANNVDFKFAEWGVTDFEM 868
```

```
RESULT 6
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. 6867026
```

RESULT 8
US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
FILE REFERENCE: BO 43388

```

RESULT 10
US-09-008-172-2
/ Sequence 2, Application US/09008172
/ Patent No. 6127602
/ GENERAL INFORMATION:
/ APPLICANT: Nichols, Scott E.
/ TITLE OF INVENTION: Substitutes for Modified Starch and
/ TITLE OF INVENTION: Latexes in Paper Manufacture
/ FILE REFERENCE: 0358D
/ CURRENT APPLICATION NUMBER: US/09/008,172
/ CURRENT FILING DATE: 1998-01-16
/ EARLIER APPLICATION NUMBER: 08/482,711
/ EARLIER FILING DATE: 1995-06-07
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 2
/ LENGTH: 1430
/ TYPE: PRT
/

```

```
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match          61.9%; Score 65; DB 2; Length 1430;
Best Local Similarity 66.7%; Pred. No. 0.17;
Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 0;

QY 3 IANNVDKFSWGTTDFEM 20
   |||||
Db 883 IAQNVQLFKSWGVTSEFEM 900

RESULT 11
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match          61.9%; Score 65; DB 2; Length 1430;
Best Local Similarity 66.7%; Pred. No. 0.17;
Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 0;

QY 3 IANNVDKFSWGTTDFEM 20
   |||||
Db 883 IAQNVQLFKSWGVTSEFEM 900

RESULT 12
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172

Query Match          53.3%; Score 56; DB 1; Length 1577;
Best Local Similarity 61.1%; Pred. No. 4.7;
Matches 11; Conservative 0; Mismatches 7; Indels 7; Gaps 0;

QY 2 VIANNVDKFSWGTTDFE 19
   |||||
Db 958 VIAQNAKLFKENGITSFE 975

; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match          61.9%; Score 65; DB 2; Length 1430;
Best Local Similarity 66.7%; Pred. No. 0.17;
Matches 12; Conservative 1; Mismatches 5; Indels 5; Gaps 0;

QY 3 IANNVDKFSWGTTDFEM 20
   |||||
Db 883 IAQNVQLFKSWGVTSEFEM 900

RESULT 13
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
; STREET: Level 8, 168 Walker Street
; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,824
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1577 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
US-08-793-824-2

Query Match          53.3%; Score 56; DB 1; Length 1577;
Best Local Similarity 61.1%; Pred. No. 4.7;
Matches 11; Conservative 0; Mismatches 7; Indels 7; Gaps 0;

QY 2 VIANNVDKFSWGTTDFE 19
   |||||
Db 958 VIAQNAKLFKENGITSFE 975
```

RESULT 14
US-09-604-957-7
; Sequence 7, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-7

Query Match 50.5%; Score 53; DB 2; Length 535;
Best Local Similarity 60.0%; Pred. No. 4;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
Db 451 VRIAQNADFFASLGFTSFEM 470

RESULT 15
US-09-995-749A-13
; Sequence 13, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-13

Query Match 50.5%; Score 53; DB 2; Length 535;
Best Local Similarity 60.0%; Pred. No. 4;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
Db 451 VRIAQNADFFASLGFTSFEM 470

Search completed: February 11, 2006, 19:46:20
Job time : 1.43789 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 20:39:23 ; Search time 1.79012 Seconds
(without alignments)
4668.162 Million cell updates/sec

Title: US-10-797-821-43

Perfect score: 105

Sequence: 1 VVIANNVDFKFSWGITDFEM 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_Main:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	20	5	US-10-797-821-43
2	105	100.0	1590	4	US-10-383-930-37
3	105	100.0	1590	5	US-10-797-821-37
4	90	85.7	1375	3	US-09-740-274-4
5	90	85.7	1375	4	US-10-383-930-35
6	90	85.7	1375	5	US-10-797-821-35
7	90	85.7	1475	3	US-09-740-274-2
8	90	85.7	1475	4	US-10-383-930-34
9	90	85.7	1475	5	US-10-797-821-34
10	71	67.6	1518	4	US-10-383-930-40
11	71	67.6	1518	5	US-10-797-821-40
12	69	65.7	1365	4	US-10-383-930-39
13	69	65.7	1365	5	US-10-797-821-39
14	68	64.8	522	3	US-09-995-749A-11
15	65	61.9	545	3	US-09-995-749A-10
16	65	61.9	1430	3	US-09-740-274-6
17	65	61.9	1430	4	US-10-383-930-36
18	65	61.9	1430	5	US-10-797-821-36
19	65	61.9	1777	5	US-10-484-218-12
20	63	60.0	1595	5	US-10-484-218-20
21	54	51.4	1006	5	US-10-484-218-22
22	53	50.5	535	3	US-09-995-749A-13
23	53	50.5	1781	3	US-09-995-749A-2
24	50	47.6	1554	4	US-10-383-930-38
25	50	47.6	1554	5	US-10-797-821-38
26	50	47.6	1771	5	US-10-484-218-14
27	49	46.7	493	5	US-10-472-928-444

28	49	46.7	661	5	US-10-617-320-3700	Sequence 3700, Ap
29	49	46.7	1497	5	US-10-484-218-18	Sequence 18, Appl
30	47.5	45.2	144	4	US-10-767-701-57788	Sequence 57788, A
31	47	44.8	354	4	US-10-282-122A-46798	Sequence 46798, A
32	47	44.8	787	5	US-10-484-218-16	Sequence 16, Appl
33	46.5	44.3	479	4	US-10-437-963-189793	Sequence 189793, A
34	46	43.8	227	4	US-10-424-599-273936	Sequence 273936, A
35	46	43.8	342	4	US-10-369-493-19927	Sequence 19927, A
36	46	43.8	402	4	US-10-724-972A-5214	Sequence 5214, Ap
37	45	42.9	399	5	US-10-732-923-10174	Sequence 10174, A
38	45	42.9	584	3	US-09-995-749A-12	Sequence 12, Appl
39	45	42.9	816	4	US-10-282-122A-52967	Sequence 52967, A
40	45	42.9	890	4	US-10-437-963-171378	Sequence 171378, A
41	45	42.9	2057	4	US-10-417-280A-2	Sequence 2, Appli
42	44.5	42.4	415	5	US-10-739-930-8402	Sequence 8402, Ap
43	44.5	42.4	449	4	US-10-425-114-46598	Sequence 46598, A
44	44.5	42.4	449	4	US-10-425-114-61364	Sequence 61364, A
45	44.5	42.4	468	4	US-10-425-115-212787	Sequence 212787, A

ALIGNMENTS

RESULT 1

US-10-797-821-43

; Sequence 43, Application US/10797821

; Publication No. US20050031633A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens

; FILE REFERENCE: 25669-020

; CURRENT APPLICATION NUMBER: US/10/797,821

; PRIOR FILING DATE: 2004-03-09

; PRIOR APPLICATION NUMBER: 10/383,930

; PRIOR FILING DATE: 2003-03-07

; PRIOR APPLICATION NUMBER: 60/363,209

; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: 60/402,483

; PRIOR FILING DATE: 2002-08-08

; PRIOR APPLICATION NUMBER: 09/290,049

; PRIOR FILING DATE: 1999-04-12

; PRIOR APPLICATION NUMBER: 60/081,550

; PRIOR FILING DATE: 1998-04-13

; PRIOR APPLICATION NUMBER: 60/115,142

; PRIOR FILING DATE: 1999-01-08

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 43

; LENGTH: 20

; TYPE: PRT

; ORGANISM: Streptococcus sobrinus

US-10-797-821-43

Query Match 100.0%; Score 105; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGITDFEM 20

DB 1 VVIANNVDFKFSWGITDFEM 20

RESULT 2

US-10-383-930-37

; Sequence 37, Application US/10383930

; Publication No. US20040127400A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A

; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

; FILE REFERENCE: 25669-018

; CURRENT APPLICATION NUMBER: US/10/383,930


```
;
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37
```

```
Query Match 100.0%; Score 105; DB 4; Length 1590;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDFVSWGITDFEM 20
|||||
DB 847 VVIANNVDFVSWGITDFEM 866
```

```
RESULT 3
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37
```

```
Query Match 100.0%; Score 105; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDFVSWGITDFEM 20
|||||
DB 847 VVIANNVDFVSWGITDFEM 866
```

```
RESULT 4
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
```

```
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4
```

```
Query Match 85.7%; Score 90; DB 3; Length 1375;
Best Local Similarity 80.0%; Pred. No. 5.4e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDFVSWGITDFEM 20
|||||
DB 876 VVIANNVDFVSWGITDFEM 895
```

```
RESULT 5
US-10-383-930-35
; Sequence 35, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-35
```

```
Query Match 85.7%; Score 90; DB 4; Length 1375;
Best Local Similarity 80.0%; Pred. No. 5.4e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 VVIANNVDFVSWGITDFEM 20
|||||
DB 876 VVIANNVDFVSWGITDFEM 895
```

```
RESULT 6
US-10-797-821-35
; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
```

; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-35

Query Match 85.7%; Score 90; DB 5; Length 1375;
Best Local Similarity 80.0%; Pred. No. 5.4e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20
||| ||||| ||:||||
Db 876 VVIANNVDKFAEWGVTDFEM 895

RESULT 7
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 85.7%; Score 90; DB 3; Length 1475;
Best Local Similarity 80.0%; Pred. No. 5.8e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20
||| ||||| ||:||||
Db 849 VVIANNVDKFAEWGVTDFEM 868

RESULT 8
US-10-383-930-34
; Sequence 34, Application US/10383930

; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-34

Query Match 85.7%; Score 90; DB 4; Length 1475;
Best Local Similarity 80.0%; Pred. No. 5.8e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20
||| ||||| ||:||||
Db 849 VVIANNVDKFAEWGVTDFEM 868

RESULT 9
US-10-797-821-34
; Sequence 34, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-34

Query Match 85.7%; Score 90; DB 5; Length 1475;
Best Local Similarity 80.0%; Pred. No. 5.8e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSVSGITDFEM 20
||| ||||| ||:||||
Db 849 VVIANNVDKFAEWGVTDFEM 868

RESULT 10
US-10-383-930-40
; Sequence 40, Application US/10383930
; Publication No. US20040127400A1

GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-383-930-40

Query Match 67.6%; Score 71; DB 4; Length 1518;
Best Local Similarity 72.2%; Pred. No. 0.054;
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 IANNVDKFSWGITDFEM 20
||| ||| ||| ||| |||
Db 922 IAEVDLFLKSWGVTSEFM 939

RESULT 11
US-10-797-821-40
; Sequence 40, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Streptococcus salivarius
US-10-797-821-40

Query Match 67.6%; Score 71; DB 5; Length 1518;
Best Local Similarity 72.2%; Pred. No. 0.054;
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 IANNVDKFSWGITDFEM 20
||| ||| ||| ||| |||
Db 922 IAEVDLFLKSWGVTSEFM 939

RESULT 12
US-10-383-930-39
; Sequence 39, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-383-930-39

GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-383-930-39

Query Match 65.7%; Score 69; DB 4; Length 1365;
Best Local Similarity 73.7%; Pred. No. 0.099;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIANNVDKFSWGITDFEM 20
||| ||| ||| ||| |||
Db 837 VIAKNGDLFKSWGITQFEM 855

RESULT 13
US-10-797-821-39
; Sequence 39, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-797-821-39

Query Match 65.7%; Score 69; DB 5; Length 1365;
Best Local Similarity 73.7%; Pred. No. 0.099;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIANNVDKFSWGITDFEM 20
||| ||| ||| ||| |||
Db 837 VIAKNGDLFKSWGITQFEM 855

RESULT 14
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. US2002015568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA

```

; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; . LENGTH: 522
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-11

```

```

Query Match          64.8%; Score 68; DB 3; Length 522;
Best Local Similarity 55.0%; Pred. No. 0.052;
Matches 11; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1  VTIANNVDKFSVSGITDFEM 20
Db      438  VTIANNVDKFSVSGITDFEM 457

```

```

RESULT 15
US-09-995-749A-10
; Sequence 10, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; . LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-995-749A-10

```

```

Query Match          61.9%; Score 65; DB 3; Length 545;
Best Local Similarity 66.7%; Pred. No. 0.16;
Matches 12; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      3  IANNVDKFSVSGITDFEM 20
Db      463  IANNVDKFSVSGITDFEM 480

```

Search completed: February 11, 2006, 20:57:30
Job time : 2.79012 secs

THIS PAGE BLANK (USPTO)


```
; APPLICANT: Blum, David
; APPLICANT: Esteghlalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 564462007901
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)...(23)
US-10-517-939-34
```

```
Query Match 39.0%; Score 41; DB 6; Length 604;
Best Local Similarity 62.5%; Pred. No. 39;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 10 FVSWGITD 17
|:|:|:|
Db 299 FITWGVTD 306
```

```
RESULT 3
US-10-517-939-134 1
; Sequence 134, Application US/10517939
; Publication No. US20060003433A1
; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esteghlalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 134
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-517-939-134
```

```
Query Match 38.1%; Score 40; DB 6; Length 336;
Best Local Similarity 53.3%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 5 NNVDKFSWGITDPE 19
|:|:|:|
Db 322 NHQKESFWGIVDFE 336
```

```
RESULT 4
US-10-519-390-20
; Sequence 20, Application US/10519390
; Publication No. US2006000872A1
; GENERAL INFORMATION:
; APPLICANT: MEDEXGEN Inc.
; APPLICANT: CHUNG, Yong-Hoon
; APPLICANT: LEE, Hak-sup
; APPLICANT: YI, Ki-Wan
; APPLICANT: KIM, Jae-Youn
; APPLICANT: HEO, Youn-Hwa
; TITLE OF INVENTION: A method of improving efficacy of biological response-modifying
; TITLE OF INVENTION: proteins and the example mutants
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/519,390
; CURRENT FILING DATE: 2004-12-23
; PRIOR APPLICATION NUMBER: KR10-2003-0051846
; PRIOR FILING DATE: 2003-07-26
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 20
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: LIP: 41st, 52nd, 67th, 70th, 156th or 180th Phe is replaced by
; OTHER INFORMATION: Val.
US-10-519-390-20
```

```
Query Match 37.1%; Score 39; DB 6; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 5 NNVDKFSWGITDF 18
|:|:|:|
Db 54 NNLDKLCGPNVTDF 67
```

```
RESULT 5
US-11-176-830-213
; Sequence 213, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Dritanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 213
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Genbank CAA32147
; DATABASE ENTRY DATE: 1995-03-22
US-11-176-830-213
```

```
Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```


QY 5 NNVDKFVSWGITDF 18
||:|
Db 54 NNLDKLCGPNVTDF 67

RESULT 6

US-11-176-830-729
; Sequence 729, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding Nu
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 729
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-729

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 NNVDKFVSWGITDF 18
||:|
Db 54 NNLDKLCGPNVTDF 67

RESULT 7

US-11-176-830-730
; Sequence 730, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding Nu
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 730
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-730

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 NNVDKFVSWGITDF 18
||:|
Db 54 NNLDKLCGPNVTDF 67

RESULT 8

US-11-176-830-731
; Sequence 731, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 731
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-731

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 NNVDKFVSWGITDF 18
||:|
Db 54 NNLDKLCGPNVTDF 67

RESULT 9

US-11-176-830-732
; Sequence 732, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 732
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-732

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;

```
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSVSGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 10
US-11-176-830-733
; Sequence 733, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; PRIOR FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 733
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-733

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSVSGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 11
US-11-176-830-734
; Sequence 734, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; PRIOR FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 734
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-734

Query Match 37.1%; Score 39; DB 7; Length 180;
```

```
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSVSGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 12
US-11-176-830-735
; Sequence 735, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; PRIOR FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 735
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-735

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 5 NNVDKFSVSGITDF 18
Db 54 NNLDKLCGPNVTDF 67

RESULT 13
US-11-176-830-736
; Sequence 736, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drittanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding N
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; PRIOR FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 736
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-736
```

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 NNVDKFKVSWGITDF 18
||:|
Db 54 NNLDKLCGPNVTDF 67

RESULT 14

US-11-176-830-737
; Sequence 737, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drihtanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding Nu
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 737
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-737

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 NNVDKFKVSWGITDF 18
||:|
Db 54 NNLDKLCGPNVTDF 67

RESULT 15

US-11-176-830-738
; Sequence 738, Application US/11176830
; Publication No. US20060020116A1
; GENERAL INFORMATION:
; APPLICANT: Gantier, Rene
; APPLICANT: Guyon, Thierry
; APPLICANT: Drihtanti, Lila
; APPLICANT: Vega, Manuel
; TITLE OF INVENTION: Rational Evolution of Cytokines for Higher Stability, Encoding Nu
; TITLE OF INVENTION: Acid Molecules and Related Applications
; FILE REFERENCE: 17109-012002 (922B)
; CURRENT APPLICATION NUMBER: US/11/176,830
; CURRENT FILING DATE: 2005-07-06
; PRIOR APPLICATION NUMBER: 10/658,834
; PRIOR FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: 60/457,135
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/409,898
; PRIOR FILING DATE: 2002-09-09
; NUMBER OF SEQ ID NOS: 1306
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 738
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-176-830-738

Query Match 37.1%; Score 39; DB 7; Length 180;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 NNVDKFKVSWGITDF 18
||:|
Db 54 NNLDKLCGPNVTDF 67

Search completed: February 11, 2006, 20:59:03
Job time : 1.11381 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2006, 19:15:50 ; Search time 0.412809 Seconds
(without alignments)
4661.567 Million cell updates/sec

Title: US-10-797-821-43

Perfect score: 105

Sequence: 1 VVIANNVDKFSWGITDFEM 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 80.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	1592	2 A38175	glucosyltransferas
2	90	85.7	1375	2 JTO345	dextranucrase (EC
3	90	85.7	1475	2 B33135	gtfB protein precu
4	71	67.6	1518	2 A44811	glucosyltransferas
5	69	65.7	1365	2 A41483	glucosyltransferas
6	65	61.9	1431	2 A45866	dextranucrase (EC
7	64	61.0	1290	2 JCS473	dextranucrase (EC
8	63	60.0	1599	2 S22737	glucosyltransferas
9	58	55.2	1449	2 T30857	glucosyltransferas
10	58	55.2	1449	2 T30552	glucosyltransferas
11	56	53.3	1577	2 T30858	glucosyltransferas
12	55	52.4	382	2 AH2361	endo-1,4-beta-xyla
13	53	50.5	1508	2 T31098	probable dextranu
14	51	48.6	659	2 G97906	transcription anti
15	49	46.7	493	2 B95036	transcription regu
16	47	44.8	178	2 S44624	C50C3.5 protein -
17	47	44.8	396	2 T11709	nifs protein homol
18	46	43.8	300	2 I40425	carboxylesterase (
19	45.5	43.3	837	1 A31842	endo-1,4-beta-xyla
20	45	42.9	200	2 T47531	hypothetical prote
21	44	41.9	412	2 T42024	probable serine pr
22	44	41.9	810	2 JC4837	hypoxia-inducible
23	43.5	41.4	133	2 AE2629	apaGB protein (impo
24	43.5	41.4	133	2 C97411	apaGB protein (AJ0
25	43	41.0	116	2 F82805	conserved hypothet
26	43	41.0	167	2 H81252	NADH2 dehydrogenas
27	43	41.0	271	2 T37745	CAAX prenyl protei
28	43	41.0	295	2 I64050	probable integral
29	43	41.0	304	2 A49496	malate dehydrogena

30	43	41.0	456	2 AH2778	ABC transporter, s
31	43	41.0	456	2 D97558	hypothetical prote
32	43	41.0	813	2 JC5809	hypoxia-inducible
33	43	41.0	826	2 I38972	hypoxia-inducible
34	42.5	40.5	301	2 C84474	hypothetical prote
35	42.5	40.5	1136	2 T40355	hypothetical prote
36	42	40.0	85	2 H81049	hypothetical prote
37	42	40.0	146	2 C90454	hypothetical prote
38	42	40.0	165	2 G88473	protein F40H6.1 (i
39	42	40.0	200	2 E81314	probable periplasm
40	42	40.0	242	2 F97158	hypothetical prote
41	42	40.0	331	2 AE3372	toluene tolerance
42	42	40.0	334	2 AD1049	HfIC protein (EC 3
43	42	40.0	334	2 G86113	probable integral
44	42	40.0	334	2 G91272	probable integral
45	42	40.0	334	2 C43653	probable integral

ALIGNMENTS

RESULT 1

A38175

Glucosyltransferase precursor - Streptococcus sobrinus

C:Species: Streptococcus sobrinus

C>Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004

C/Accession: A38175

R/Abn. H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

J. Bacteriol. 173, 989-996, 1991

A>Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus

A/Reference number: A38175; MUID:91123227; PMID:1704006

A/Accession: A38175

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1592 <ABO>

A/Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:G217032; PIDN:BAA14241.1; PID

F:1093-1112/Domain: cpl repeat homology <CP1>

F:1222-1241/Domain: cpl repeat homology <CP2>

F:1287-1306/Domain: cpl repeat homology <CP3>

F:1330-1351/Domain: cpl repeat homology <CP4>

F:1352-1371/Domain: cpl repeat homology <CP5>

F:1402-1420/Domain: cpl repeat homology <CP6>

F:1465-1484/Domain: cpl repeat homology <CP7>

F:1513-1532/Domain: cpl repeat homology <CP8>

Query Match 100.0%; Score 105; DB 2; Length 1592;

Best Local Similarity 100.0%; Pred. No. 5.8e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDKFSWGITDFEM 20

|||||

Db 847 VVIANNVDKFSWGITDFEM 866

RESULT 2

JTO345

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

N/Alternate names: sucrose 6-glucosyltransferase

C:Species: Streptococcus mutans

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004

C/Accession: JTO345; C33135

R/Ueda, S.; Shiroza, T.; Kuramitsu, H.K.

Gene 69, 101-109, 1988

A>Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.

A/Reference number: JTO345; MUID:89137980; PMID:2976010

A/Accession: JTO345

A/Molecule type: DNA

A/Residues: 1-1375 <UED>

A/Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515

A/Experimental source: GS-5

R/Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A>Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: C33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <SHI>
A:Cross-references: UNIPARC:UPI000014E25D; GB:M17361
C:Genetics:
A:Gene: gtfc
C:Function:

A:Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans
C:Keywords: duplication; glycosyltransferase; hexosyltransferase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-1375/Product: glucosyltransferase #status predicted <MAT>
F:1126-1145/Domain: cpl repeat homology <CP1>
F:1253-1272/Domain: cpl repeat homology <CP2>
F:1318-1337/Domain: cpl repeat homology <CP3>

Query Match 85.7%; Score 90; DB 2; Length 1375;
Best Local Similarity 80.0%; Pred. No. 1.1e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGITDFEM 20
||||| ||||| ||:|||||
Db 876 VVIANNVDFKFAENGVTDFEM 895

RESULT 3

B33135
gtfB protein precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Jul-2004
C:Accession: B33135; A33128
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
J. Bacteriol. 169, 4263-4270, 1987
A:Title: Sequence analysis of the gtfB gene from Streptococcus mutans.
A:Reference number: A33135; MUID:87308013; PMID:3040685
A:Accession: B33135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1475 <SHI>
A:Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN:
R:Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
submitted to the Protein Sequence Database, September 1990
A:Reference number: A33128
A:Accession: A33128
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-171,173-641,'N',643-1475 <SH2>
A:Cross-references: UNIPARC:UPI000017AC5E
A:Experimental source: strain GS-5
F:1096-1115/Domain: cpl repeat homology <CP1>
F:1224-1243/Domain: cpl repeat homology <CP2>
F:1289-1308/Domain: cpl repeat homology <CP3>
F:1354-1373/Domain: cpl repeat homology <CP4>
F:1419-1438/Domain: cpl repeat homology <CP5>

Query Match 85.7%; Score 90; DB 2; Length 1475;
Best Local Similarity 80.0%; Pred. No. 1.2e-05;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGITDFEM 20
||||| ||||| ||:|||||
Db 849 VVIANNVDFKFAENGVTDFEM 868

RESULT 4

A4811
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
C:Species: Streptococcus salivarius
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A44811; S22726; S28809
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
J. Gen. Microbiol. 137, 2577-2593, 1991

A:Title: Molecular characterization of a cluster of at least two glucosyltransferase ge
A:Reference number: A44811; MUID:92148377; PMID:1838391
A:Accession: A44811
A:Molecule type: DNA
A:Residues: 1-1518 <GIF>
A:Cross-references: UNIPROT:Q00600; UNIPARC:UPI000000BEF31; EMBL:Z11873; NID:g47526; PIDN:
A:Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIP:81052)
C:Genetics:
A:Gene: gtff
C:Function:

A:Description: glycosyltransferase; hexosyltransferase
C:Keywords: cpl repeat homology <CP4>
F:1307-1326/Domain: cpl repeat homology <CP4>

Query Match 67.6%; Score 71; DB 2; Length 1518;
Best Local Similarity 72.2%; Pred. No. 0.011;
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 IANNVDFKFSWGITDFEM 20
||||| ||||| ||:|||||
Db 922 IANNVDFKFSWGVTSEFEM 939

RESULT 5

A41483
glucosyltransferase (EC 2.4.1.-) gtfs precursor - Streptococcus sobrinus
C:Species: Streptococcus sobrinus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Mar-2004
C:Accession: A41483
R:Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
Infect. Immun. 58, 2452-2458, 1990
A:Title: Analysis of the Streptococcus downei gtfs gene, which specifies a glucosyltran
A:Reference number: A41483; MUID:90316665; PMID:2142479
A:Accession: A41483
A:Molecule type: DNA
A:Residues: 1-1365 <GIL>
A:Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:g153652; PIDN:AAA36898.1; PID
C:Genetics:
A:Gene: gtfs
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 65.7%; Score 69; DB 2; Length 1365;
Best Local Similarity 73.7%; Pred. No. 0.02;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIANVDFKFSWGITDFEM 20
||||| ||||| ||:|||||
Db 837 VIAKNGDLFKSWGITQFEM 855

RESULT 6

A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C:Accession: A45866
R:Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A:Title: Nucleotide sequence of the Streptococcus mutans gtfd gene encoding the glucosyl
A:Reference number: A45866; MUID:91100958; PMID:2148600
A:Accession: A45866
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1431 <HON>
A:Cross-references: UNIPARC:UPI000017AC5C; GB:M29296
C:Keywords: glycosyltransferase; hexosyltransferase
F:181-201/Domain: cpl repeat homology <CP1>
F:1127-1146/Domain: cpl repeat homology <CP2>
F:1192-1211/Domain: cpl repeat homology <CP3>
F:1257-1276/Domain: cpl repeat homology <CP4>
F:1277-1297/Domain: cpl repeat homology <CP5>
F:1321-1340/Domain: cpl repeat homology <CP8>
F:1341-1361/Domain: cpl repeat homology <CP6>
F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match	↑	61.9%	Score 55;	DB 2;	Length 1431;
Best Local Similarity		66.7%	Pred. No. 0.087;		
Matches	12;	Conservative	1;	Mismatches	5; Indels 0; Gaps 0;
Qy	3	IANNVDRKFSVSGITDFEM 20			
Db	883	IAQNVQLFKSGVTSFEM 900			
RESULT 7					
JC5473					
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides					
C:Species: Leuconostoc mesenteroides					
C:Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004					
C:Accession: JC5473					
R:Monchois, V.; Willemet, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.					
Gene 182, 23-32, 1996					
A:Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc mesenteroides					
A:Reference number: JC5473; MUID:97136686; PMID:8982063					
A:Accession: JC5473					
A:Status: nucleic acid sequence not shown; translation not shown					
A:Molecule type: DNA					
A:Residues: 1-1290 <MON>					
A:Cross-references: UNIPROT:Q48756; UNIPARC:UPI0000017ABDF; GB:U38181					
C:Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto					
C:Genetics:					
A:Gene: dsrA					
C:Keywords: glycosyltransferase; hexosyltransferase					
F:78-870/Domain: catalytic #status predicted <CAT>					
F:922-1290/Domain: glucan-binding #status predicted <GCB>					
Query Match		61.0%	Score 64;	DB 2;	Length 1290;
Best Local Similarity		60.0%	Pred. No. 0.11;		
Matches	12;	Conservative	2;	Mismatches	6; Indels 0; Gaps 0;
Qy	1	VVIANNVDRKFSVSGITDFEM 20			
Db	686	IKIAQNVNLFKDWGITSPFEM 705			
RESULT 8					
S22737					
glucosyltransferase (EC 2.4.1.1) S - Streptococcus salivarius					
C:Species: Streptococcus salivarius					
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004					
C:Accession: S22737; S28810; B44811; S22727					
R:Jacques, N.					
submitted to the EMBL Data Library, March 1992					
A:Reference number: S22726					
A:Accession: S22737					
A:Molecule type: DNA					
A:Residues: 1-1599 <JAC>					
A:Cross-references: UNIPROT:Q00599; UNIPARC:UPI000000BEF34; EMBL:Z11872; MID:g47530; PIDN					
A:Experimental source: ATCC 25975					
R:Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.					
J. Gen. Microbiol. 137, 2577-2593, 1991					
A:Title: Molecular characterization of a cluster of at least two glucosyltransferase gen					
A:Reference number: A44811; MUID:92148377; PMID:1838391					
A:Accession: S28810					
A:Molecule type: DNA					
A:Residues: 1-51 <GIF>					
A:Cross-references: UNIPARC:UPI000017027C; EMBL:Z11873					
C:Genetics:					
A:Gene: gtfK					
C:Keywords: glycosyltransferase; hexosyltransferase					
F:1456-1475/Domain: cpl repeat homology <CPR>					
Query Match		60.0%	Score 63;	DB 2;	Length 1599;
Best Local Similarity		63.2%	Pred. No. 0.2;		
Matches	12;	Conservative	2;	Mismatches	5; Indels 0; Gaps 0;
Qy	2	VVIANNVDRKFSVSGITDFEM 20			
		↑			

A:Gene: qtfm

Query Match 53.3%; Score 56; DB 2; Length 1577;
Best Local Similarity 61.1%; Pred. No. 2.4;
Matches 11; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 VIANNVDKFSWGITDFE 19
||| | | | | | |
Db 958 VIAONAKLFKEWGITSTFE 975

RESULT 12

AH2361
 endo-1,4-beta-xylanase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 31-Dec-2004
 C:Accession: AH2361
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH2361
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-382 <KUR>
 A:Cross-references: UNIPROT:O8YNW3; UNIPARC:UPI000000CEE23; GB:BA0000019; PIDN:BAE76147.1;
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr4448
 C:Superfamily: xylanase; Streptomyces endo-1,4-beta-xylanase A homology

Query Match	52.4%	Score 55;	DB 2;	Length 382;
Best Local Similarity	64.3%	Pred. No. 0.77;		
Matches	9:	Conservative	2:	Mismatches
			3:	Indels
				Gaps 0:

Qy 4 ANNVDFVSWGTD 17
||| : |||
Db 315 ANGVDTITWGTGTD 328

RESULT 13

T311098
probable dextranucrase (EC 2.4.1.5), extracellular - *Leuconostoc mesenteroides*
C/Species: *Leuconostoc mesenteroides*
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C/Accession: T31098
R/Monchois, V.; Renaud-gimeon, M.; Monsan, P.; Willemot, R.M.
FEMS Microbiol. Lett. 199, 307-315, 1998
A/Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DSR)
A/Reference number: 220981; MUID:98164374; PMID:9503626
A/Accession: T31098
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1508 <MON>
A/Cross-references: UNIPROT:O52224; UNIPARC:UPI00000B569B; EMBL:AF030129; NID:g2766611;
A/Experimental source: strain NRRL B-1299
C/Genetics:
A/Gene: dsrB
C/Function:
A/Description: produces dextran composed only of alpha (1-6) glucosidic bonds
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 50.5%; Score 53; DB 2; Length 1508;
 Best Local Similarity 52.6%; Pred. No. 6.8;
 Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy	2	VIANNVDFVSWGITDFEM	20
Db	927	VIKNGOLFCKDWGITSFOL	94

RESULT 14

G97906
transcription antiterminator BglG family bglG [imported] - Streptococcus pneumoniae (strain G97906)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: G97906
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; Feyer, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Meyer, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Ballido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MOID:21429245; PMID:1154234
A:Accession: G97906
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-659 <KUR>
A:Cross-references: UNIPROT:Q8DR86; UNIPARC:UPI00000E33FC; GB:AE007317; PIDN:AAK9083.1.1; C:Genetics: bglG
C:Gene: bglG

Query Match	48.6%;	Score 51;	DB 2;	Length 659;
Best Local Similarity	50.0%;	Pred. No. 5.8;		
Matches	9:	Conservative	4:	Mismatches
				5: Indels
				Gaps 0:

Qy 3 IANVDKFVSWGITDFEM 20
| | : | | : | | : | :
pb 371 ILANMDKFELEWSISDDDEI 388

RESULT 15

B95036
transcription regulator, probable [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: B95036
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Elsen, J.A.; Read, T.D.; Peterson, S.; Hein-
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95036
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-493 <NR>
A:Cross-references: UNIPROT:Q97SM4; UNIPARC:UPI00000513AC; GB:AE005672; PIDN:AAK74483.1.
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0306

Query Match	46.7%	Score 49;	DB 2;	Length 493;
Best Local Similarity	50.0%	Pred. No. 8.7;		
Matches	9;	Conservative	4;	Mismatches
			5;	Indels
			0;	Gaps

Qy 3 IANNVDKFVSWGITDFEM 20
| | | | | : | | | | :
Db 371 ILANMDKFELEWTISDDEV 388

Search completed: February 11, 2006, 19:43:00
Job time : 1.41281 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: February 11, 2006, 18:57:35 ; Search time 2.69869 Seconds
(without alignments)
5228.676 Million cell updates/sec
Title: US-10-797-821-43
Perfect score: 105
Sequence: 1 VVIANNVDKFSWGITDFEM 20
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	1590	2 Q55263_9STRE	Q55263 streptococc
2	105	100.0	1590	2 Q59983_9STRE	Q59983 streptococc
3	105	100.0	1592	1 GTF2_STRDO	P27470 streptococc
4	105	100.0	1597	1 GTF1_STRDO	P11001 streptococc
5	90	85.7	1591	2 Q8VUH3_STRMU	Q8VUH3 streptococc
6	90	85.7	1455	1 GTF3_STRMU	P13470 streptococc
7	90	85.7	1476	1 GTFB_STRMU	P08987 streptococc
8	72	68.6	1338	2 Q9WXJ4_9STRE	Q9WXJ4 streptococc
9	72	68.6	2835	2 Q8G9Q2_LEUME	Q8G9Q2 leuconostoc
10	71	67.6	1518	2 Q00600_STRSL	Q00600 streptococc
11	69	65.7	1016	2 Q9LCJ7_LEUME	Q9LCJ7 leuconostoc
12	69	65.7	1365	1 GTF5_STRDO	P29336 streptococc
13	69	65.7	1454	2 Q6XA94_LEUME	Q6XA94 leuconostoc
14	68	64.8	1522	2 Q6TXV4_LEUME	Q6TXV4 leuconostoc
15	68	64.8	1527	2 Q8KRE1_LEUME	Q8KRE1 leuconostoc
16	68	64.8	1527	2 Q9ZAR4_LEUME	Q9ZAR4 leuconostoc
17	65	61.9	1462	1 GTFD_STRMU	P49331 streptococc
18	65	61.9	1772	2 Q5SBN3_LACRE	Q5SBN3 lactobacill
19	64	61.0	1290	2 Q48756_LEUME	Q48756 leuconostoc
20	64	61.0	1506	2 Q56CX8_9STRE	Q56CX8 streptococc
21	64	61.0	1512	2 Q9WXJ5_9STRE	Q9WXJ5 streptococc
22	63	60.0	1575	2 Q9LCH3_STROR	Q9LCH3 streptococc
23	63	60.0	1577	2 Q54178_STRGN	Q54178 streptococc
24	63	60.0	1595	2 Q5SBN3_LACSK	Q5SBN3 lactobacill
25	63	60.0	1599	2 Q00599_STRSL	Q00599 streptococc
26	58	55.2	1449	2 Q68542_STRSL	Q68542 streptococc
27	58	55.2	1449	2 Q55264_STRSL	Q55264 streptococc
28	56	53.3	1577	2 Q55265_STRSL	Q55265 streptococc
29	55	52.4	382	2 Q8YNW3_ANASP	Q8YNW3 anabaena sp
30	54	51.4	1463	2 Q5SBN6_LACFE	Q5SBN6 lactobacill
31	53	50.5	1330	2 Q84CN4_LEUME	Q84CN4 leuconostoc

32	53	50.5	1477	2 Q9L466_LEUME	Q9L466 leuconostoc
33	53	50.5	1508	2 Q9EZH5_LEUME	Q9EZH5 leuconostoc
34	53	50.5	1508	2 Q52224_LEUME	Q52224 leuconostoc
35	53	50.5	1781	2 Q5SBL9_LACRE	Q5SBL9 lactobacill
36	53	50.5	1781	2 Q4JCS4_LACRE	Q4JCS4 lactobacill
37	53	50.5	1781	2 Q4JLC7_LACRE	Q4JLC7 lactobacill
38	51	48.6	659	2 Q8DR86_STRR6	Q8DR86 streptococc
39	51	48.6	798	2 Q9RKFO_STRCO	Q9RKFO streptomyc
40	51	48.6	1319	2 Q5JEU0_PYRKO	Q5JEU0 pyrococcus
41	50	47.6	119	2 Q6UB60_9STRA	Q6UB60 hyaloperono
42	50	47.6	1554	2 Q8KZL5_9STRE	Q8KZL5 streptococc
43	50	47.6	1772	2 Q5SBN0_LACRE	Q5SBN0 lactobacill
44	49	46.7	119	2 Q7T4W3_9ENTO	Q7T4W3 human echov
45	49	46.7	119	2 Q7T4W4_9ENTO	Q7T4W4 human echov

ALIGNMENTS

RESULT 1
Q55263_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263; 9STRE PRELIMINARY; PRT; 1590 AA.
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DE 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE GTF-1.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus;
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase produced from Streptococcus sobrinus ATCC 33478.";
RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D63570; BAA09792.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; 1GVM.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6B4FD43 CRC64;
Query Match 100.0%; Score 105; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VVIANNVDKFSWGITDFEM 20
Db 847 VVIANNVDKFSWGITDFEM 866
RESULT 2
Q59983_9STRE PRELIMINARY; PRT; 1590 AA.
ID Q59983_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q59983;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)

DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfI;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]_TaxID=1310;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OMZ176;
RX MEDLINE=91122498; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus";
RL DNA Seq. 4:19-27(1993).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; LHXC.
DR GO; GO:0047849; F: dextranucrase activity; IEA.
DR GO; GO:0016757; F: transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P: glucan biosynthesis; IEA.
DR InterPro; IPR002479; G: CW binding.
DR InterPro; IPR003318; G: Glyco_hydro_70.
DR Pfam; PF01473; C: CW_binding_1; 2.
DR Pfam; PF02324; G: Glyco_hydro_70; 1.
KW Glycosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 Glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 105; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
|||||
DB 847 VVIANNVDFVSWGITDFEM 866

RESULT 3
GTF2_STRDO
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=6715 / Serotype G;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus Glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.

CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
CC fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 16 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.

CC EMBL; D90213; BAA14241.1; -; Genomic_DNA.
DR HSSP; P06653; LGVM.
DR InterPro; IPR002479; Cell wall bd put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; C: CW_binding_1; 3.
DR Pfam; PF02324; G: Glyco_hydro_70; 1.
KW Dental caries; Glucosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1592 Glucosyltransferase-I.
FT REPEAT 1093 1142 1.
FT REPEAT 1158 1207 2.
FT REPEAT 1222 1272 3.
FT REPEAT 1287 1337 4.
FT REPEAT 1402 1451 5.
FT REPEAT 1514 1563 6.
FT REPEAT 1577 1592 7 (incomplete).
FT REGION 39 1044 Catalytic (approximate).
FT REGION 1093 1592 7 X tandem repeats.
FT REGION 1093 1592 Glucan-binding (approximate).
SQ SEQUENCE 1592 AA; 176168 MW; BC0A66D079351ECF CRC64;

Query Match 100.0%; Score 105; DB 1; Length 1592;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFVSWGITDFEM 20
|||||
DB 847 VVIANNVDFVSWGITDFEM 866

RESULT 4
GTF1_STRDO
ID GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfI;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
RT sobrinus MFE28";
RL J. Bacteriol. 169:4271-4278(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-

```

CC fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 19 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL: M17391; AAC63063.1; -; Genomic DNA.
CC InterPro: IPR002479; Cell wall bd put.
CC InterPro: IPR003318; Glyco_hydro_70.
CC Pfam: PF01473; CW_binding_1; 4.
CC DR PFam; PF02324; Glyco_hydro_70; 1.
CC DR Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
CC KW SIGNAL 38 Potential.
CC FT CHAIN 35 1597 Glycosyltransferase-I.
CC FT REPEAT 1099 1132 A repeat.
CC FT REPEAT 1163 1213 AC repeat.
CC FT REPEAT 1227 1277 AC repeat.
CC FT REPEAT 1292 1342 AC repeat.
CC FT REPEAT 1352 1399 B repeat.
CC FT REPEAT 1406 1455 AC repeat.
CC FT REPEAT 1465 1512 B repeat.
CC FT REPEAT 1519 1568 AC repeat.
CC FT REPEAT 1582 1597 A repeat (incomplete).
CC FT REGION 39 1050 Catalytic (approximate).
CC FT REGION 1099 1597 1-25 A, 2 B and 5 AC repeats.
CC FT REGION 1099 1597 Glucan-binding (approximate).
CC SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 100.0%; Score 105; DB 1; Length 1597;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGTTDFEM 20
DB 853 VVIANNVDFKFSWGTTDFEM 872
|||||

```

```

DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
FT NON_TER 1 1
SQ SEQUENCE 591 AA; 67094 MW; 0933DCB4421DAF30 CRC64;

Query Match 85.7%; Score 90; DB 2; Length 591;
Best Local Similarity 80.0%; Pred. No. 9.9e-06;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVIANNVDFKFSWGTTDFEM 20
DB 12 VVIANNVDFKFSWGTTDFEM 31
|||||

```

RESULT 6

GTFC_STRMU STANDARD; PRT; 1455 AA.

AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)
DE (Dextranucrase) (sucrose 6-glucosyltransferase).
GN Name=gTfC; OrderedLocNames=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]

NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT4148 / Serotype c;
RX MEDLINE=98231843; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]

NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [4]

NUCLEOTIDE SEQUENCE OF 1-349.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [5]

-!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
CC fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes

water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

-!- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.

-!- SIMILARITY: Contains 5 cell wall binding repeats.

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

EMBL; M2054; AAA88592.1; -; Genomic_DNA.
 DR EMBL; D8652; BAA26102.1; -; Genomic_DNA.
 DR EMBL; D8655; BAA26106.1; -; Genomic_DNA.
 DR EMBL; D8658; BAA26110.1; -; Genomic_DNA.
 DR EMBL; D8661; BAA26114.1; -; Genomic_DNA.
 DR EMBL; D8978; BAA26120.1; -; Genomic_DNA.
 DR EMBL; AB014940; AA058706.1; -; Genomic_DNA.
 DR EMBL; M17361; AA088589.1; -; Genomic_DNA.
 DR EMBL; J0345; J0345.
 DR HSP; P06653; IH8G.
 DR InterPro; IPR002479; Cell_wall_bd_put.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 2.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 DR Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
 KW Transferase.

SIGNAL 1 34
 FT CHAIN 39 1455 Glucosyltransferase-SI.
 FT REPEAT 1126 1159 A repeat.
 FT REPEAT 1169 1200 A repeat.
 FT REPEAT 1227 1238 C repeat.
 FT REPEAT 1253 1303 AC repeat.
 FT REPEAT 1318 1330 A repeat (incomplete).
 FT REGION 35 1050 Catalytic (approximate).
 FT REGION 1126 1455 2.4 A, 1 C and 1 AC repeats.
 FT REGION 1126 1455 Glucan-binding (approximate).
 FT VARIANT 21 21 V -> I (in strain GS-5).
 FT VARIANT 81 81 P -> L (in strain MT4239).
 FT VARIANT 106 106 D -> V (in strain GS-5).
 FT VARIANT 116 116 S -> A (in strain GS-5 and strain MT4467).
 FT VARIANT 126 126 A -> T (in strain GS-5).
 FT VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239 and strain MT4467).
 FT VARIANT 256 256 A -> V (in strain GS-5 and strain MT4467).
 FT VARIANT 425 425 R -> N (in strain MT4251).
 FT VARIANT 519 519 Y -> D (in strain MT4245 and strain MT4251).
 FT VARIANT 538 538 R -> K (in strain MT4245 and strain MT4251).
 FT VARIANT 545 545 Y -> F (in strain MT4245 and strain MT4251).
 FT VARIANT 597 597 N -> D (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
 FT VARIANT 600 600 R -> K (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
 FT VARIANT 601 601 A -> T (in strain GS-5).
 FT VARIANT 614 614 M -> T (in strain GS-5).
 FT VARIANT 727 727 T -> I (in strain MT8148).
 FT VARIANT 734 734 A -> V (in strain MT8148).
 FT VARIANT 964 964 L -> F (in strain MT4239).
 FT VARIANT 1113 1113 N -> Y (in strain MT4239).
 FT VARIANT 1118 1118 I -> V (in strain MT4239).
 FT VARIANT 1204 1204 I -> V (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
 FT VARIANT 1208 1208 V -> I (in strain MT8148).
 FT VARIANT 1292 1294 DGH -> NGY (in strain GS-5, strain MT4467 and strain MT8148).
 FT VARIANT 1305 1369 Missing (in strain MT4245).
 FT VARIANT 1326 1326 I -> V (in strain MT8148).
 FT VARIANT 1331 1331 T -> A (in strain GS-5, strain MT4239,

FT VARIANT 1377 1377 strain MT4467 and strain MT8148).
 FT VARIANT 1398 1398 R -> K (in strain MT8148).
 FT VARIANT 1424 1424 V -> I (in strain MT8148).
 FT VARIANT 1439 1439 D -> N (in strain MT4239).
 FT VARIANT 1439 1439 V -> I (in strain MT4239 and strain MT8148).
 FT VARIANT 1444 1444 S -> P (in strain MT8148).
 FT CONFLICT 1337 1455 QRLYFKSGVQAKGELITERKRIKYDPSNGEVNRVVR
 FT TSSGNNYYFGNDGYALIGHVVEGRVRRVDFDNGVRYASHD
 FT QRHWYDYRRDFGRGSSAVFRHSRNGFDNFRFF ->
 FT HASILSLMVFRLRESSLSQSKVVSNTMLIPMKFVIM
 FT (in Ref. 1).
 SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A9A4FEC86 CRC64;
 Query Match 85.7%; Score 90; DB 1; Length 1455;
 Best Local Similarity 80.0%; Pred. No. 2.5e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VVIANNVDKFSWGVTDFEM 20
 Db 876 VVIANNVDKFAEWGVTDFEM 895

RESULT 7
 ID GTFB_STRMU STANDARD; PRT; 1476 AA.
 AC P08987; O69381; O69384; O69387; O69390; O69396;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
 DE (Sucrose 6-glucosyltransferase).
 GN Name=gtfB; OrderedLocusNames=SMU.1004;
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1309;
 RN [1]
 RC NUCLEOTIDE SEQUENCE.
 RP STRAIN=GS-5;
 RX MEDLINE=87308013; PubMed=3040685;
 RA Shiroza T., Ueda S., Kuramitsu H.K.;
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
 RL J. Bacteriol. 169:4263-4270(1987).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
 RX MT4467 / Serotype e, and MT8148 / Serotype c;
 RX MEDLINE=982331643; PubMed=9570124;
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
 RA Kimura S., Hamada S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 Streptococcus mutans.";
 RL FEMS Microbiol. Lett. 161:331-336(1998).
 RN [3]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
 RA Carson M.B., Prineaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
 RA Ferretti J.J.;
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
 pathogen.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to
 play a key role in the development of the dental plaque because of
 their ability to adhere to smooth surfaces and mediate the
 aggregation of bacterial cells and food debris.
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
 fructose + (1,6-alpha-D-glucosyl) (n+1).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha

1,3-linked glucose and some 1,6 linkages), GTF-8 synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

-1- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.

-1- SIMILARITY: Contains 10 cell wall binding repeats.

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

EMBL; M17361; AAA88588.1; -; Genomic DNA.
 EMBL; D88653; BAA26101.1; -; Genomic DNA.
 EMBL; D88654; BAA26105.1; -; Genomic DNA.
 EMBL; D88657; BAA26109.1; -; Genomic DNA.
 EMBL; D88660; BAA26113.1; -; Genomic DNA.
 EMBL; D89977; BAA26119.1; -; Genomic DNA.
 EMBL; AE014940; AAN58705.1; -; Genomic DNA.
 PIR; B33135; B33135.
 DR HSSP; P06653; IH8G.
 DR InterPro; IPR002479; Cell wall bd put.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 4.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.

FT SIGNAL 1 34 Potential.
 FT CHAIN 35 1476 Glucosyltransferase-I.
 FT REPEAT 1097 1130 A repeat.
 FT REPEAT 1161 1210 1.
 FT REPEAT 1225 1275 2.
 FT REPEAT 1290 1340 3.
 FT REPEAT 1355 1405 4.
 FT REPEAT 1420 1470 5.
 FT REGION 35 1051 Catalytic (approximate).
 FT REGION 1097 1476 Glucan-binding (approximate).
 FT REGION 1161 1470 S X tandem repeats.
 FT VARIANT 62 62 S -> T (in strain MT4239).
 FT VARIANT 65 65 T -> I (in strain GS-5).
 FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
 FT VARIANT 78 78 Q -> P (in strain MT4251).
 FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
 FT VARIANT 89 89 S -> P (in strain MT4251).
 FT VARIANT 168 168 K -> N (in strain MT4251).
 FT VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and strain MT8148).
 FT VARIANT 399 399 N -> R (in strain MT4239).
 FT VARIANT 474 474 I -> T (in strain MT4239).
 FT VARIANT 512 512 K -> R (in strain MT8148).
 FT VARIANT 519 519 F -> Y (in strain MT8148).
 FT VARIANT 701 701 T -> I (in strain MT8148).
 FT VARIANT 708 708 A -> V (in strain MT8148).
 FT VARIANT 938 938 F -> L (in strain MT8148).
 FT VARIANT 952 957 FGKPYE -> YGIPVA (in strain GS-5, strain MT4239 and strain MT4467).
 FT VARIANT 963 964 SV -> NT (in strain GS-5, strain MT4239 and strain MT4467).
 FT VARIANT 968 970 ADS -> VDG (in strain GS-5, strain MT4239 and strain MT4467).
 FT VARIANT 1086 1086 A -> T (in strain MT4239).
 FT VARIANT 1158 1158 S -> N (in strain MT4239).
 FT VARIANT 1163 1163 H -> Y (in strain MT4251).
 FT VARIANT 1168 1168 E -> K (in strain MT8148).
 FT VARIANT 1182 1182 Y -> C (in strain MT8148).
 FT VARIANT 1234 1234 A -> P (in strain MT4239).
 FT VARIANT 1263 1263 R -> H (in strain GS-5 and strain MT4467).
 FT VARIANT 1263 1263 R -> P (in strain MT8148).

FT VARIANT 1264 1264 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
 FT VARIANT 1272 1272 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
 FT VARIANT 1329 1329 H -> Y (in strain GS-5 and strain MT4467).
 FT VARIANT 1394 1394 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
 FT VARIANT 1402 1402 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
 FT VARIANT 1459 1459 Y -> H (in strain MT4467).
 FT CONFLICT 570 570 R -> A (in Ref. 1).
 FT CONFLICT 800 817 ADQDVRAASTAPSTDGK -> LKMFALRLARPHQOMA (in Ref. 1).
 FT CONFLICT 1310 1310 H -> L (in Ref. 1).
 SQ SEQUENCE 1476 AA; 165847 MW; 9C6E09F731B4CBFC CRC64;
 Query Match 85.7%; Score 90; DB 1; Length 1476;
 Best Local Similarity 80.0%; Pred. No. 2.6e-05;
 Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VIANNVDFVSWGITDPEM 20
 ||||| ||||| ||||| |||||
 Db 850 VVIAKNVDFVSWGITDPEM 869

RESULT 8
 Q9WKJ4_9STRE PRELIMINARY; PRT; 1338 AA.
 AC Q9WKJ4;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE GTF-S.
 GN Name-gtfS;
 OS Streptococcus criceti.
 OG Plasmid pAM1.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1333;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=HS-6;
 RA Inoue M., Fukui K., Miyagi A.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB026123; BAA77236.1; -; Genomic_DNA.
 DR HSSP; P06653; IH8G.
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.
 DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 2.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Plasmid.
 SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;
 Query Match 68.6%; Score 72; DB 2; Length 1338;
 Best Local Similarity 73.7%; Pred. No. 0.02;
 Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 VIANNVDFVSWGITDPEM 20
 ||||| ||||| ||||| |||||
 Db 809 VVIAKNVDFVSWGITDPEM 827

RESULT 9
 Q8G9Q2_LEUME PRELIMINARY; PRT; 2835 AA.
 ID Q8G9Q2;
 AC Q8G9Q2;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Dextranucrase (EC 2.4.1.5) (Fragment).
 GN Name=dexE;

CC primer glucan unlike GTP-1.
 CC -1- MISCELLANEOUS: Synthesizes water-soluble glucans (alpha 1,6-
 CC glucose).
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
 CC -1- SIMILARITY: Contains 10 cell wall binding repeats.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL: M30943; AAA26898.1; -; Genomic_DNA.
 CC HSSP: P06653; 1H8G.
 CC InterPro: IPR002479; Cell_wall_bd_put.
 CC InterPro: IPR003318; Glyco_hydro_70.
 CC Pfam: PF01473; CW_binding_1; 2.
 CC Pfam: PF02324; Glyco_hydro_70; 1.
 KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
 FT SIGNAL 1 36
 FT CHAIN 37 1365 Glucosyltransferase-S.
 FT REPEAT 157 177 Cell wall binding 1.
 FT REPEAT 178 197 Cell wall binding 2.
 FT REPEAT 1062 1082 Cell wall binding 3.
 FT REPEAT 1082 1102 Cell wall binding 4.
 FT REPEAT 1150 1169 Cell wall binding 5.
 FT REPEAT 1170 1190 Cell wall binding 6.
 FT REPEAT 1225 1243 Cell wall binding 7.
 FT REPEAT 1289 1308 Cell wall binding 8.
 FT REPEAT 1309 1328 Cell wall binding 9.
 FT REPEAT 1331 1352 Cell wall binding 10.
 FT REGION 198 1061 Catalytic (approximate).
 FT SEQUENCE 1365 AA; 151591 MW; 167296BSA2E8C476 CRC64;
 Query Match 65.7%; Score 69; DB 1; Length 1365;
 Best Local Similarity 73.7%; Pred. No. 0.061; Indels 5; Gaps 0;
 Matches 14; Conservative 0; Mismatches 5; Indels 5; Gaps 0;
 QY 2 VIANNVDFKFSWGITDFEM 20
 DB 837 VIANNVDFKFSWGITDFEM 855
 RESULT 13
 Q69A94 LEUME
 ID Q69A94 LEUME PRELIMINARY; PRT; 1454 AA.
 AC Q69A94
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Dextranase (EC 2.4.1.5).
 GN Name=derP;
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=IBT-PQ;
 RA Fernandez-Vazquez J.L., Lopez-Munguia A., Olvera C.;
 RT "Molecular characterization of a dextranase gene from Leuconostoc
 RT mesenteroides IBT-PQ isolated from pulque."
 RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY504865; AAS79426.1; -; Genomic DNA.
 DR GO: GO:0047849; F:dextranase activity; IEA.
 DR GO: GO:0016757; F:transfructose activity; IEA.
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.
 DR InterPro: IPR002479; CW binding.
 DR InterPro: IPR003318; Glyco_hydro_70.
 DR Pfam: PF01473; CW_binding_1; 1.
 DR Pfam: PF02324; Glyco_hydro_70; 1.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;

Query Match 65.7%; Score 69; DB 2; Length 1454;
 Best Local Similarity 68.4%; Pred. No. 0.066;
 Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VIANNVDFKFSWGITDFE 19
 DB 887 VQIANVTDLYKSWGINTFE 905
 RESULT 14
 Q6TXV4 LEUME
 ID Q6TXV4 LEUME PRELIMINARY; PRT; 1522 AA.
 AC Q6TXV4
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Dextranase (EC 2.4.1.5).
 GN Name=derX;
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=L0309;
 RA Wang J.-H., Shao Y.-C., Teng D., Yang Y.-L., Zhang F.;
 RT "Cloning and sequence analysis of gene coding for dextranase from
 RT Leuconostoc mesenteroides L0309."
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AV743959; AAQ98615.2; -; Genomic DNA.
 DR GO: GO:0047849; F:dextranase activity; IEA.
 DR GO: GO:0016757; F:transfructose activity; IEA.
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.
 DR InterPro: IPR002479; CW binding.
 DR InterPro: IPR003318; Glyco_hydro_70.
 DR Pfam: PF01473; CW_binding_1; 2.
 DR Pfam: PF02324; Glyco_hydro_70; 1.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 1522 AA; 169103 MW; 01BCC15468B913AB CRC64;
 Query Match 64.8%; Score 68; DB 2; Length 1522;
 Best Local Similarity 55.0%; Pred. No. 0.1;
 Matches 11; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 VIANNVDFKFSWGITDFEM 20
 DB 940 VVIAQNAQDFKQGVTSFQL 959
 RESULT 15
 Q8KRE1 LEUME
 ID Q8KRE1 LEUME PRELIMINARY; PRT; 1527 AA.
 AC Q8KRE1
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Dextranase (EC 2.4.1.5).
 GN Name=derD;
 OS Leuconostoc mesenteroides.
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
 OX NCBI_TaxID=1245;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC MEDLINE=22573396; PubMed=12686639; DOI=10.1099/mic.0.26029-0;
 RA Neubauer H., Bauche A., Mollet B.;
 RT "Molecular characterization and expression analysis of the
 RT dextranase DsrD of Leuconostoc mesenteroides Lcc4 in homologous
 RT and heterologous Lactococcus lactis cultures."
 RL Microbiology 149:973-982(2003).
 DR EMBL: AY017384; AAG61158.1; -; Genomic_DNA.
 DR HSSP: P06653; 1H8G.
 DR GO: GO:0047849; F:dextranase activity; IEA.
 DR GO: GO:0016757; F:transfructose activity; IEA.
 DR GO: GO:0009250; P:glucan biosynthesis; IEA.

DR InterPro; IPR002479; CW_binding.
 DR InterPro; IPR003318; Glyco_hydro_70.
 DR Pfam; PF01473; CW_binding_1; 2.
 DR Pfam; PF02324; Glyco_hydro_70; 1.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 1527 AA; 169835 MW; F9D0DE220BD89668 CRC64;
 Query Match 64.8%; Score 68; DB 2; Length 1527;
 Best Local Similarity 55.0%; Pred. NO. 0.1;
 Matches 11; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 1 VVIANNVDFVSVSGITDEEM 20
 Db 945 VVIQNADQFKQWGVTSFQL 964

Search completed: February 11, 2006, 19:39:14
 Job time : 2.69869 secs